AD-A174 757

SCHEMA-BASED MODEL OF INFORMATION PROCESSING FOR SITUATION ASSESSMENT

David Noble
Deborah Boehm-Davis
Carla Grosz

Engineering Research Associates
8618 Westwood Center Drive
Vienna, Virginia 22180

OCTOBER 1986

Sponsored By:

Engineering Psychology Programs
Office of Naval Research
Contract N00014-84-C-0484
Work Unit Number NR 649-005

86 12 04 044

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER	2. GOVT ACCESSION NO. ADA 174757	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) A Schema-Based Model of Information Processing For Situation Assessment		5. TYPE OF REPORT & PERIOD COVERED Interim Report
		6. PERFORMING ORG. REPORT NUMBER R-622-86
7. AUTHOR(s) David F. Noble, Deborah Boehm-Davis		8. CONTRACT OR GRANT NUMBER(s) N00014-84-C-0484
9. PERFORMING ORGANIZATION NAME AND ADDRESS Engineering Research Associates 8618 Westwood Center Drive Vienna, Virginia 22180		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR 649-D05
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research 800 North Quincy Street Arlington, Virginia 22217		12. REPORT DATE September 1986
		13. NUMBER OF PAGES 74
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

This document has been approved for public release and sale; its distribution is unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

schema
Information processing
situtation assessment

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

A model of human information processing that links the properties of presented information to situation assessments based on this information is described, and data from experiments testing the major model attributes are presented. The model is based on the assumption that humans assess situations by determining the extent to which features in an observed situation match features of previously encountered reference situations that have been encoded as schema within semantic memory. The model involves seven processing steps. The first six are: schema activation at task initiation, object classification

DD FORM 1473, 1 JAN 73 EDITION OF 1 NOV 65 IS OBSOLETE

S/N 0102-LF-014-6601

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

substitution of objects not usually processed by the schema with functional equivalents that can be processed, scaling of features using schema criteria curves, weighting of features, and combining features using a weighted geometric mean of scaled features. The seventh step is an iteration of the last three. The experimental data strongly support the proposed model as a descriptive model of human information processing. The model has value, as well, for the design of information presentation aids.

A-1

R-622-86

SCHEMA-BASED MODEL OF INFORMATION PROCESSING FOR SITUATION ASSESSMENT

David Noble
Deborah Boehm-Davis
Carla Grosz

Engineering Research Associates
8618 Westwood Center Drive
Vienna, Virginia 22180

October 1986

Sponsored By:

Engineering Psychology Programs
Office of Naval Research
Contract N00014-84-C-0484
Work Unit Number NR 649-005

TABLE OF CONTENTS

TABLE OF CONTENTS
(Continued)

TABLE OF CONTENTS
(Continued)

R-622-86

A SCHEMA-BASED MODEL OF INFORMATION PROCESSING FOR SITUATION ASSESSMENT

The research reported here tests a model of human information processing that links the properties of presented information to the situation assessments based on this information (Noble, 1985). The model focuses on the kind of information processing required to support judgmental processes that are based primarily on experience and situation recognition. These include processes that may underlie "intuitive" decision making. This model is based primarily on schema* theory, though it also draws from other cognitive psychology notions.

Schema are memory structures used for information processing that enable people to use their experience to recognize and interpret situations, understand language and stories, make decisions, and solve problems. A well-known type of schema is the script (Bower, Black and Turner, 1979; Pryor, 1985) Scripts are time-event models of familiar experiences. The events, which partition the script into major scenes, may be scripts themselves. Schema, in general, need not be time or event oriented. They are characterized by variables, a hierarchy of embedding, and varying levels of abstraction which "attempt to represent knowledge in the kind of flexible way which reflects human tolerance for vagueness, imprecision, and quasi-inconsistencies" (Rumelhart, 1977). As recognition devices their "processing is aimed at the evaluation of their goodness of fit to the data being processed" (Rumelhart, 1980).

* We use schema, rather than schemata, as the plural.

Several researchers have shown that people use schema to understand natural language and stories (van Dijk and Kintch, 1983; Rumelhart, 1981; Thornkyke, 1977). Thorndyke showed that there exist schema that define a standard structure for stories. Stories with this structure were easier to understand than those with a different structure. Van Dijk and Kintch's model of language comprehension explains how multiple schema interacting within a linked hierarchy enables people to understand words, sentences, and paragraphs. Rumelhart showed that stories seem to be understood in terms of explanatory schema evoked by key words.

Schema have also been shown to guide actual judgment, behavior and decision making in cognitive, social, and clinical psychology (Abelson, 1981). They have been shown to be useful in several cases where problem solving is based on the ability to use methods that worked previously in similar situations. These cases have included understanding and solving arithmetic word problems (Kintsch, 1985), solving algebra problems based on their propositional structure (Mayer, 1982), and finding promising solution strategies for geometry and maze problems based on their surface features (Lewis, 1985). They have also been shown to account for differences between expert and novice approaches to solving physics problems (Larkin, 1983), and to account for some flawed heuristics and biases associated with human information processing (Tversky, 1980; Kahneman, 1973; Tversky, 1983). In their papers, Tversky and Kahneman showed that people seem to try to establish schema that can account for observed data, and then to use schema for reasoning. A classic paper on chess expertize (Chase and Simon, 1973), though not explicitly a "schema" paper, also shows that expertize can be based upon the ability to recognize chess patterns associated with previously learned good moves.

The knowledge contained in schema can be applied to a particular problem only if that problem's relevance to a schema can be recognized. Since schema enable particular instances to be recognized as belonging to a class of instances, schema can be regarded to be partly classification devices (Abelson, 1981). The present research adopts a probabilistic view of classification (Smith, 1981), in which an object or concept is classified when enough of its weighted features match the set of features associated with the category. The features can be at several levels of abstraction (Tversky, 1984; Larkin, 1983) They may be physical parts, properties such as symmetry or color, and functions (Gati, 1984). It is not presently understood how people are able to recognize the features used in classification. One possible mechanism could be based on a hierarchy of similarity assessments (Tversky, 1977) at different levels of aggregation and abstraction (Rumelhart, 1980).

Schema can support reasoning by analogy by helping people to recognize that a particular situation is related to a class of situations. In reasoning by analogy, methods proven to work for one class of problems are applied to a new class (diSessa, 1983).

It is not yet understood how schema are formed (Rumelhart, 1977), but since schema represent the results of experience, schema must somehow be generalized from a sequence of past experiences. One model (Hayes-Roth, 1977; Elio, 1981), based on feature powersets of examplars, proposes that each time a new experience is encountered which is similar to one for which a schema exists, the elements of the property set of the new experience augment in memory those property sets which have been previously stored. Our research is also based on the assumption that schema are developed from past experiences.

Accordingly, subject training, which is designed to help people quickly acquire schema, is based on presentation of examples.

The research described here builds upon many previous concepts, but is most directly related to the work of Zimmerman and Zysno (1980) which applies fuzzy set concepts to decision making. In that study, subjects rated the quality of features (fit and strength, as inferred from shape and color) of each tile to be used in a furnace, and separately rated the overall quality of that tile. Zysno and Zimmerman noted that the assessments of overall tile quality could be estimated from the geometric mean of the subjects' estimates of tile color and shape.

Schema-based information processing is described in the literature for a broad range of tasks. Some of the researchers describe general principles of schema structure and operation that are applicable to many problems. Others propose a specific structure and information processing sequence within the context of a particular problem. This paper is of the latter type. It describes a specific information processing model for situation assessment.

This model assumes that situation assessment occurs by comparing an observed situation with memory reference models for different situation types and by associating the observed situation with the reference model that it matches best. This model "represents knowledge in the kind of flexible way which reflects human tolerance for .... imprecision" (Rumelhart, 1977), thereby enabling it to accommodate inexact matches between observed and reference situations. While the model draws on the current literature and is consistent with the data in this literature, the specific model proposed here is believed to be new.

The information processing model

The human information processing model presented here is comprised of an information processing structure and information processing steps that relate the properties of presented information assessments made about that situation.

In the experiments to be described, the situations to be evaluated are "all-out attacks" or "barriers". The information processing model describes the specific steps through which presented information results in an assessment of attack or barrier quality. An example of the presented information is shown in Figure 1. This particular picture was one of several used for training. The test pictures are similar, but do not include text information about attack quality. In this figure the friendly forces (white) are positioned in the center of the picture. They are surrounded by hostile (black) ships, submarines, and aircraft. The information processing model accounts for subjects' assessments of the effectiveness of this attack in terms of the attack features.

Information processing structure

The information processing structure consists of a network of linked schema. The hypothesized structure for evaluating the quality of an all-out attack consists of four primary schema: one each for the surface, air, and submarines threats, and one for the overall attack.

Each schema (Figure 2) consists of three layers: a slot layer, a criteria layer, and an inference and action layer. Each schema can be thought of as a decision making mechanism, with each layer corresponding to a step in the decision process: the slot layer corresponds to problem formulation; the criteria layer to problem analysis; and the inference and action layer to alternative selection.

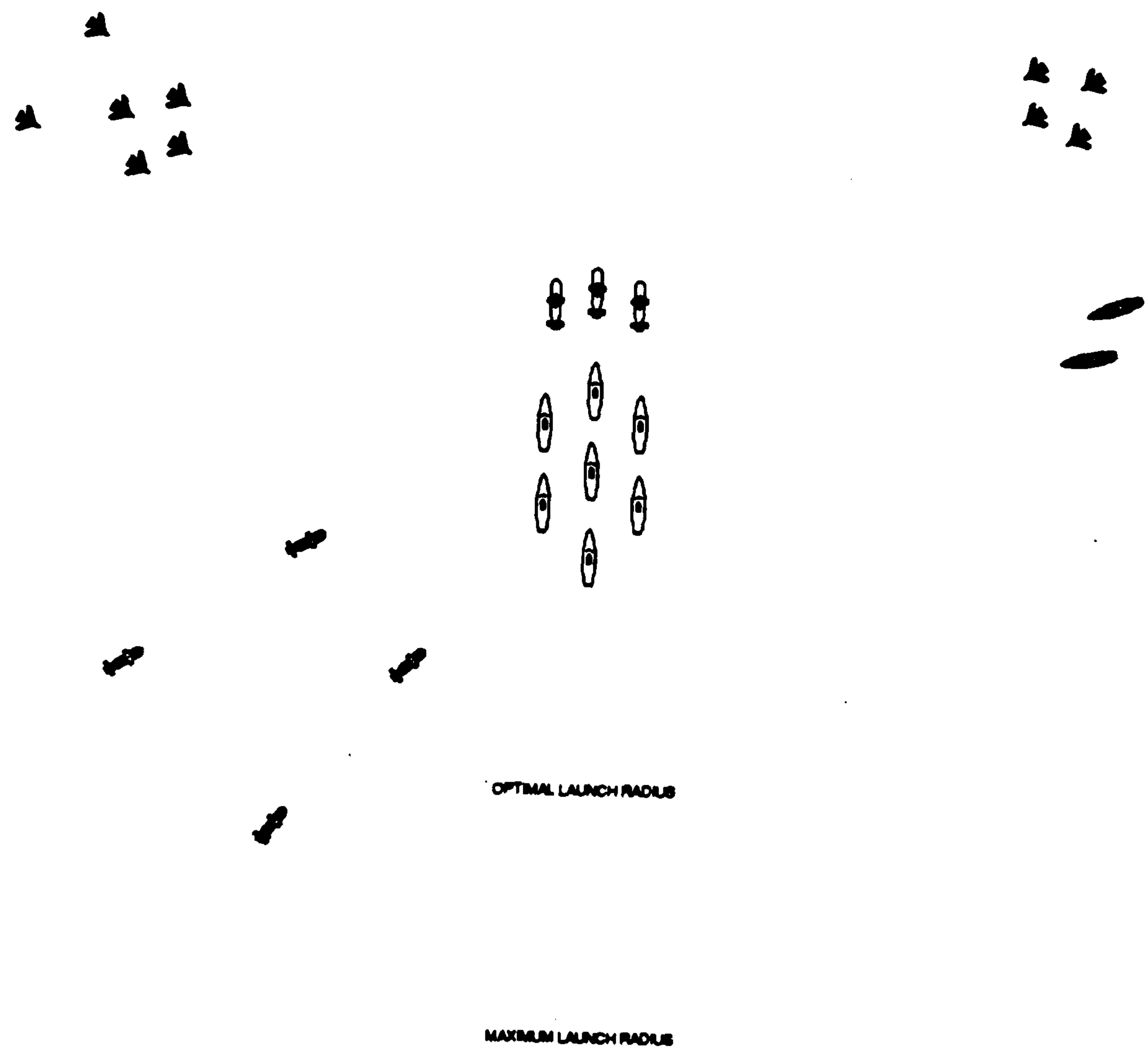


Figure 1. An example of a training picture for all-out attacks in experiment 1. Subjects were told that "Attack effectiveness is 4. The air threat is severe, but the ship and sub threats are weak. There are too few ships, and the submarines are concentrated in only a single quadrant."

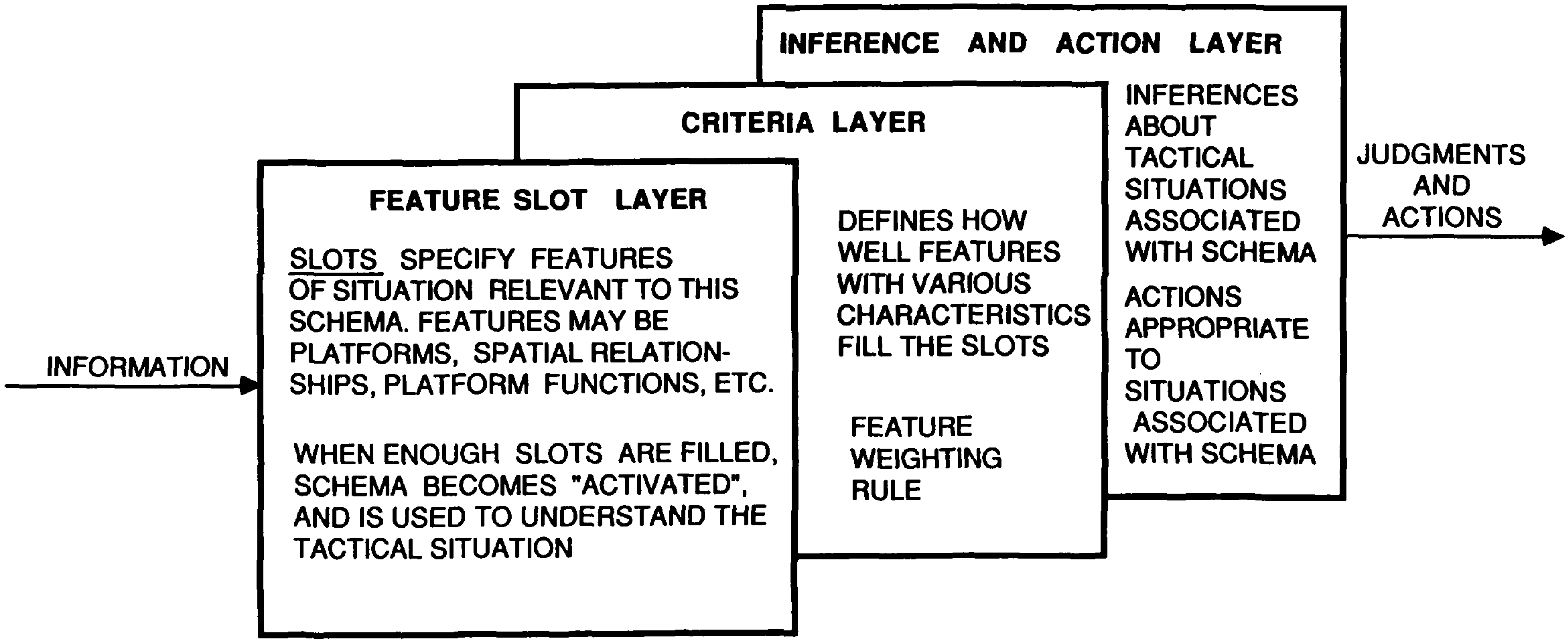


Figure 2. A Single Schema.

ONR-009

The slot layer specifies a set of slots used for identifying situation features that are relevant to the schema. Each slot specifies the physical and functional properties of potential slot fillers. The schema for the submarine threat contains a slot for the feature "many submarines" and a second slot for the feature "multi-axis threat". The schema for the overall attack contains slots for the overall surface, subsurface, and air threats.

The second layer contains data for feature assessment. Our model presents these data as feature criteria curves and weighting rules. The criteria curves convert measurable picture properties, such as the number of aircraft or the distance between ships, into subjective feature assessments such as "many aircraft" or "barrier length". In our experiments the feature assessments measure the degree to which features have characteristics consistent with a high quality hostile attack or barrier. In the schema for the surface threat, for example, the criteria curve for the feature "many ships" defines the extent to which any particular number of ships qualifies as being "many ships" in the specific context of an all-out attack. These criteria curves may be interpreted as fuzzy set membership functions in the set "many ships". In these experiments, subjective feature assessment scores range from one to ten with a score of ten indicating a feature chacteristic of a very strong attack or barrier and a score of one indicating a feature characteristic of a very weak attack or barrier. For the attacks in our first experiment, a picture with only a single ship would score about a one on the feature "many ships"; one with seven ships would score about ten on this feature.

The second layer also contains a rule for assigning feature weights. These weights reflect the relative importance of each feature in assessing overall attack or barrier effectiveness. Examples of weighting rules could

include "assign equal weights", "assign each feature a weight equal to its feature assessment score", or for a two feature schema "weight the higher feature .25 and the lower feature .75".

The third schema layer specifies the actions to be taken and inferences to be made given various levels of schema activation. Such actions and inferences are retrieved as if by a table look-up within the schema. Examples of actions to be taken are "attempt to pass through barriers of quality less than "5" (as rated by the schema)" and "do not pass through barriers rated of higher quality". This level plays no role in the present situation assessment model, and is not examined in these experiments. The inference and action level is expected to be important in models that address decision making based on situation assessment.

Schema acquistion

People are assumed to acquire schema by abstracting (usally subconsciously) a general model from specific instances. In acquiring the schema described previously, people must identify 1) a set of situation features corresponding to the schema slots, 2) a set of feature criteria curves, and 3) a feature weight assignment rule. In the present experiments, subjects acquired schema by being shown a set of examples, each associated with an attack or barrier effectiveness score and a qualitative statement about the strength of individual features. Subjects were not told the feature criteria curves or feature weighting rules, but were expected to infer these as they acquired the schema.

Figure 3 summarizes the overall schema acquisition environment in the experiments. The experimenters developed a "schema-like" situation assessment model containing feature assessment curves and a weighting rule for a specified

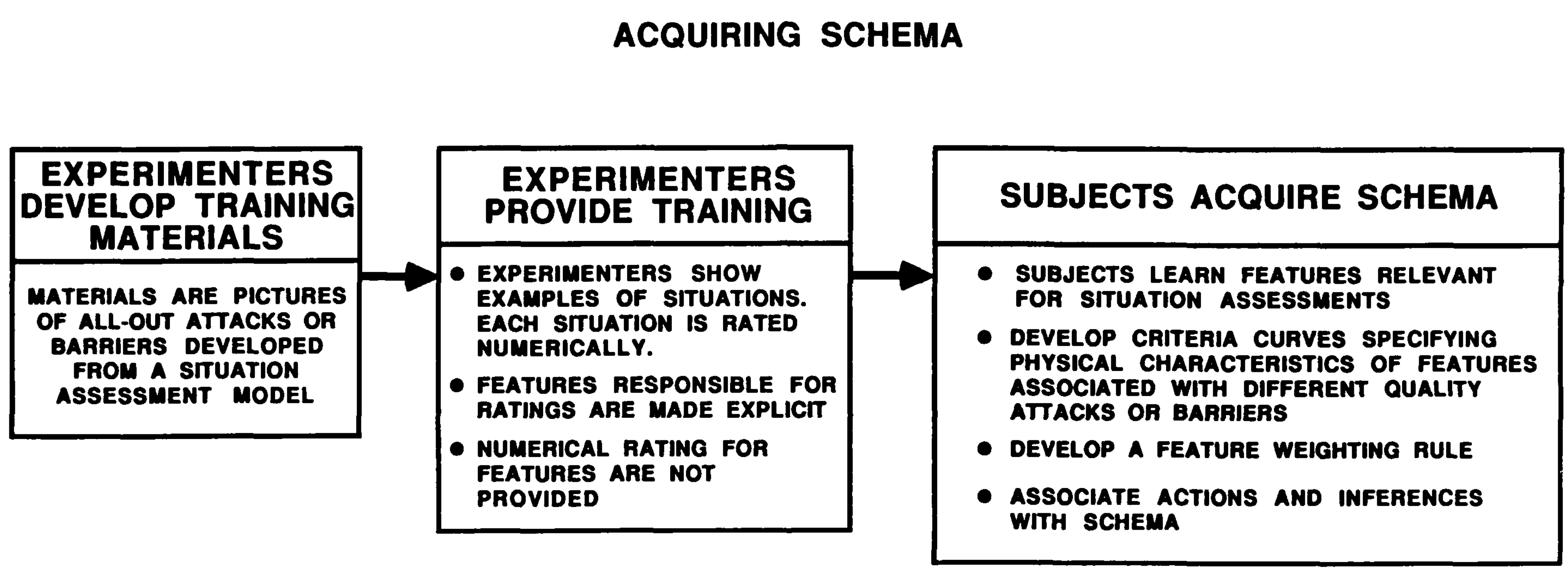


Figure 3. Material Preparation, Training and Proposed Schema Acquisition All-out Attack and Barrier Experiments

ONR 008

set of features. Subjects were then trained by being shown examples constructed from this model. It is proposed that the subjects develop the schema by abstracting the relevant features, feature criteria curves, and feature weighting rule from these examples.

Information processing steps

Figure 4 outlines seven information processing steps proposed to account for subjects' ratings of situation quality. These steps process information derived from the presented pictures by using reference data that is developed during training and stored within the network of schema.

1. Initial selection of schema. Presentation of a task will cause selection of schema related to the task. Thus, the task "evaluate the following all-out attacks" will cause schema related to attack evaluation to be made available. This step is not examined in these experiments, and is not discussed further.

2. Object classification. Subjects classify the familiar objects within each picture of an attack or barrier. It is at this point that ships are recognized as ships and not as blotches caused by a dirty copying machine. The present experiments do not examine how objects are classified, and it is not discussed further.

3. Assessment of feature relevance and functional substitution. This step examines the objects and relationships among objects classified in the previous step, attempts to find relevant features, and converts relevant features into standard physical units. This step uses information in the slot layer of the schema which specifies functional and physical properties of objects relevant to the schema. In this step all objects able to fill a schema slot are converted into the standard physical units used by the schema criteria curves for feature

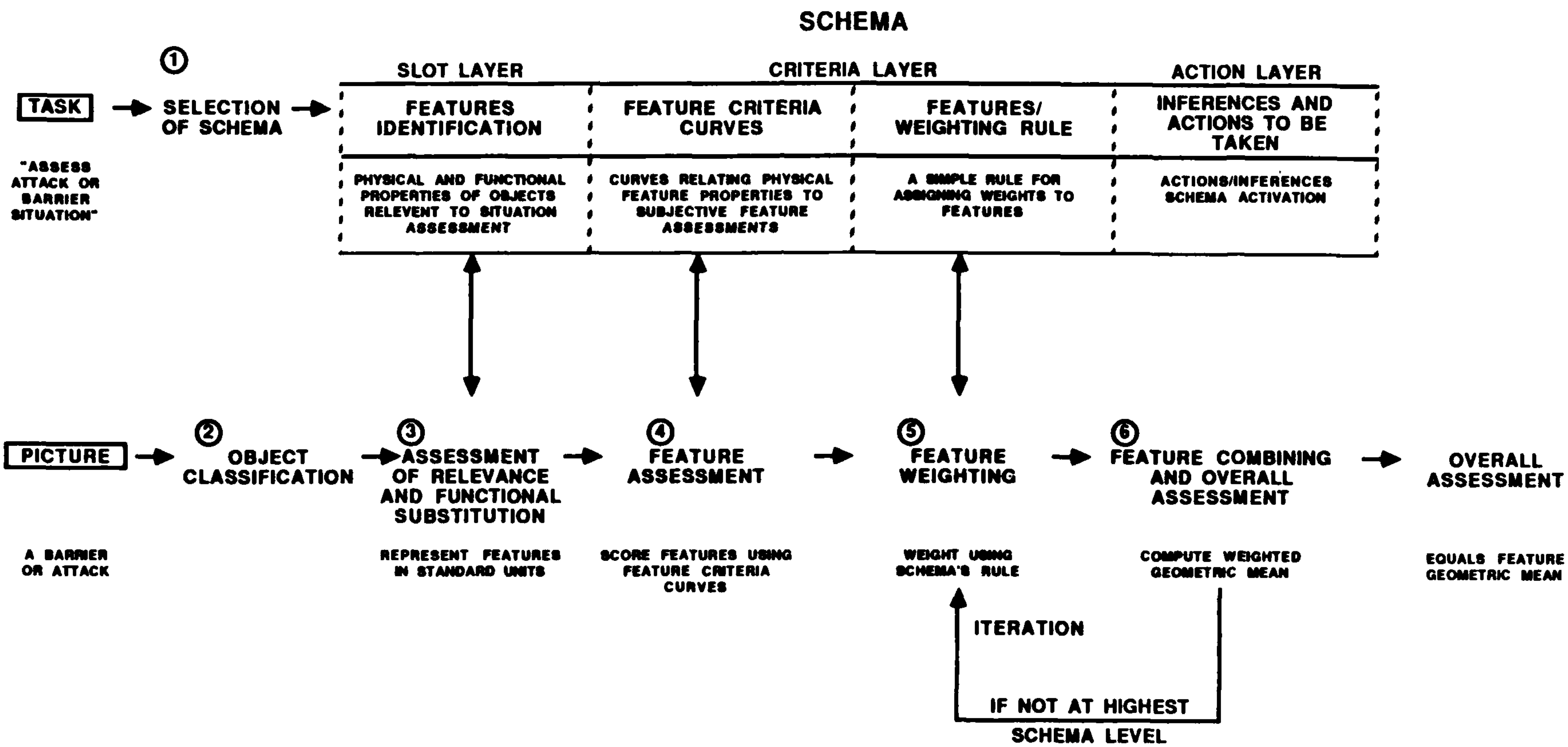


Figure 4. Use of Schema for Situation Assessment

ONR-007

assessment. In the experiment 3 barrier evaluations, islands are converted into ship equivalents in this step.

4. Feature assessment. In this step, the physical units filling the feature slots are converted into schema-specific feature assessment scores. In experiment 1, for example, the feature "many aircraft" would receive a score of about seven in any attack having eight aircraft. This and the following two steps use data stored in the criteria layer.

5. Feature weighting. Each scored feature is assigned a weight generated by the schema weight assignment rule.

6. Feature combining. Features are combined using some weighting scheme for the assessed and scored features. The geometric mean was used in the experiments reported here and it worked well, but the specific combination rule used is not important to the model. A weighted arithmetic mean would probably work about as well. The geometric rather than arithmetic mean was selected for this model because the geometric mean allows a single situation feature, which is completely inconsistent with a particular schema, to prevent that schema from being used as the situation model.

7. Iteration of steps five and six at higher schema levels. In the all-out attack example, steps four through six assessed, weighted and combined three pairs of features: many ships and ship multi-axis assessed and combined into overall ship threat; many aircraft and aircraft multi-axis assessed and combined into overall air threat; and many submarines and submarine multi-axis assessed and combined into overall submarine threat. In the iteration of steps five and six at a higher level, the overall ship, air, and submarine threats will be

weighted and combined. The result of this feature combination is the score for the attack quality. This score is the overall assessment of the attack.

Critical model issues

Experiments 1 through 3 test the the adequacy of the proposed model for mapping the connection between presented information and assessed all-out attack and barrier quality. These tests address specific information processing issues in the third through seventh steps described above; they also address the ability of subjects to acquire stable and accurate schema from a sequence of examples.

Stability, accuracy, and ease of learning of schema abstracted from examples. The training is intended to install schema for all-out attacks and barriers in a way that is consistent with the natural acquisition of schema through everyday experiences. In this training, subjects are shown ten to twelve examples of attacks or barriers. For each example, they are given a numerical rating for the quality of the barrier or attack, and told the features that contribute to its strength and weakness. They are not given numbers for feature strength, nor are they told the relative importance of the different features.

The experiments test the ease with which schema can be learned from information presented this way. They test the "accuracy" of the subjects' schema, as measured by the extent to which the subjects' assessments match the assessments predicted by the schema-like model used to develop the training pictures. In addition they test the "stability" of schema, as measured by the consistency of subjects' assessments over time. Data for ease of learning, stability, and accuracy test whether subjects attain and use a relatively permanement cognitive model for their assessments.

Ease of learning is measured by the time it takes for a subject to learn to rate a set of training pictures consistently (within one point of a standard rating for that picture). It is hypothesized that if situation evaluation is naturally mediated by data organized within schema having the structure described previously, then training materials designed to fit these structures should be easy to learn. If nearly all subjects can learn the material within two or three presentations of the set, then it seems likely that the training is taking advantage of readily developed cognitive structures.

Stability implies that people are basing their assessments on a cognitive model that is not changing through the duration of the experiment. In the all-out attack experiments subjects are presented with each test picture twice, separated by about an hour. During this hour the subjects first performed a distraction task, and then rated features in pictures of all-out attacks. Because the number of test pictures exceeds the capacity of episodic memory, and because the time interval between successive estimates of the same attacks exceeds item retention in episodic memory, stability also implies that this cognitive model is in semantic memory.

Like ease of learning and stability, accuracy implies the use of schema for the subjects' assessments. In these experiments, the training picture ratings were derived using a model that computes barrier or all-out attack quality from the feature characteristics. An accurate schema captures this model. It enables subjects to rate each picture approximately the same as the model would.

In these experiments subjects' assessments are compared to the ratings produced by the model. Accurate assessments of attacks or barriers not seen in

the training implies that these assessments are based on the model. This implication is tested directly in the barrier experiments. The ten test pictures in this experiment included five from the training set and five new pictures. If the subjects' assessments of situation quality for the new pictures were as close to the standard as were their assessments for the pictures seen previously in the training, then it may be concluded that the pictures are being evaluated using schema rather than by remembering the actual pictures presented in training.

Assessment of feature relevance and functional substitution. (Step 3 in the information processing model. The first and second steps are not examined in these experiments.) Data in schema enable relevant features to be identified and used for assessment. These data should enable people to identify and use features composed of objects physically different from, but functionally equivalent to, the objects included in the training materials. Functional substitution is this ability to use functionally equivalent objects in the schema-based assessments.

Two different mechanisms for functional substitution seem plausible. The first possibility proposed was that the functional substitution occurs very early in the processing. Objects such as ships or islands are first classified. Their relevance to the schema is then determined by comparing the physical and functional properties of the objects with the properties specified by the schema slots. If an object is determined to be relevant, then it will be used in the sitution evaluation. Since the feature criteria curves used in step four are unlikely to have been developed for objects not included in the training material, a mechanism must exist to enable existing feature criteria curves to

accommodate these objects. It is proposed that this mechanism is to substitute a functionally equivalent number of old objects for the new objects, and then to use the existing criteria curves with the old objects. Thus, in the island part of experiment 3, an island would be converted into a certain, functionally equivalent, number of ships, and then the criteria curve developed to evaluate ships is used to evaluate the effect of the islands.

A second possibility for functional substitution is that it occurs later in the evaluation process. In this case, the schema will cause each object to be evaluated according to the physical criteria developed in the training material. For example, the feature "barrier length" might be judged from the distance between the two ships at the ends of the barrier. A barrier with the two end ships far apart would be evaluated high on this feature; one with the two end ships near together would be evaluated low. If functional substitution occurs late in the process, then a barrier with two ships near one another would initially be rated the same on the "barrier length" feature whether or not there exist other objects in the picture that function as blockading ships beyond the two end ships. Thus, a short barrier completely blocking a channel inlet would score low. The late substitution alternative proposes that low scoring pictures with unusual objects would be re-evaluated. This second evaluation would not use the physical properties of the objects in the picture (ship distances, for example) but rather would use the functional properties of these objects (ability to block passage).

Both of these alternatives for functional substitution seem plausible, and each offers some advantages in information processing. The former, with early functional substitution, does not require that pictures that score poorly be reevaluated using a second set of features concerned with functionality. The

latter, which uses functional properties only when necessary, allows evaluations for most cases to be made without requiring that object functionality be considered at all.

Experiment 3 is designed to discriminate between these two alternatives. It presents pictures with objects not seen in the training pictures, and elicits overall scores and scores for "physical" (distance between ships) and "functional" (barrier is hard to go around) features. If the subjects' overall evaluation can be predicted only from the functional features, and not from the physical features, then the second alternative proposing separate functional and physical tests is supported. If the subjects' overall evaluation is predicted equally well from the physical or functional features, then the first alternative, early functional substitution, would be favored.

Feature assessment. (Step 4 in the model). In the feature assessment step, physical features, which are measureable quantities, such as the number of ships in a picture or the distance between two ships, are converted to a related subjective assessments, such as "many ships" or "barrier length". These assessments are schema specific, and actually mean "many ships for the purpose of all-out attacks", or "barrier length sufficient for barrier to be effective".

The model proposes that feature critieria curves are used to assess and score features as needed for situation assessment. If this is the case, then it is expected that the feature score would be related to an underlying physical variable in a simple monotonic way and that this relationship would not depend on the values of other features in the picture. It is also expected that the criteria curve, being schema specific, would closely reflect the training materials.

All experiments address the use of feature criteria curves. Experiments 1 and 2, however, are designed to examine these issues critically. These two experiments differ only in the number of objects in each training and test picture. Every picture in experiment 2 has 50% more hostile ships, submarines, and aircraft than the corresponding picture in experiment 1.

It is expected that if the criteria curves are determined entirely by the training pictures, then the curves from the two experiments would differ only by a 50% scaling factor. Thus, if six ships in experiment 1 receives a score of 7 for the feature "many ships", then nine ships in experiment 2 would receive a score of 7 for that feature. If this relationship is not true, then the curves must be determined both by the training material and also by general concepts related to feature evaluations. For example, nine ships in the second experiment might be scored higher than six ships in the first experiment because nine scores higher in the general category "many" than does six.

During training subjects were never given any numerical ratings for feature quality. Instead they were given only the overall picture rating and qualitative feature assessments. Because of this and the fact that there are many different ways to combine two feature scores to yield an overall rating, it is not expected that the feature criteria curves inferred by the subjects would match the criteria curves used in the model to develop the training and test materials. These experiments offer an opportunity to observe discrepencies between the model criteria curves and the curves inferred by subjects.

<u>Feature weighting (step 5 in the model)</u>. The model proposes that overall picture assessments are the weighted geometric mean of the feature scores. It is expected that in each picture some features contribute more to

the overall assessment than do other features. For example, in these experiments barrier quality depends only on two features: barrier length and barrier solidity. In the model used to develop the experiment materials overall barrier quality depended primarily on the quality of the weaker feature. For instance, a barrier that is very long but has big holes would be rated low because it is easy to pass through. On the other hand, it is expected that a barrier that is very solid but quite short would also be rated low, because it is easy to pass around.

The experiments test several issues concerned with weighting. They test whether or not people do weight features differently for different examples of all-out attack or barrier. They test whether these weightings can be derived from a simple weighting rule for all pictures corresponding to a single schema, and whether different schema have different rules.

There are two different ways for the experimenters to infer subjects' feature weights. One way is from the feature importance ratings provided by the subjects. If the feature weights are the same as the importance scores, then feature weights can be obtained directly from these ratings. Feature weights can also be attained indirectly, however, by finding a weight assignment rule that produces weighted geometric means close to the picture ratings. By comparing the weights obtained by these two methods it is possible to determine how importance ratings relate to feature weights.

Feature combination. (step 6 in model). The key prediction of this model is that the weighted geometric means of subjects' feature assessments approximate their assessments of the overall attack or barrier quality. This

prediction is tested directly in each of the three experiments. Poor correlation between these weighted means and the assessments would invalidate the model. Good correlation would suggest its utility for modeling the information processing for situation assessment.

Experiments

In this research program, three related experiments were conducted to address the issues suggested by our model of schema-based information processing. All three experiments provided data to test our hypotheses regarding the relationship between situation assessment and feature assessment, feature weighting and feature combination. These data also address the extent to which subjects are able to infer and use the model used to develop the experimental materials. In addition, each experiment provided some data on particular aspects of the model.

Experiment 1: All-Out Attack, Low Density

This experiment was designed to test several properties of schema as used for situation assessment. Specifically, this experiment provides information about criteria curves for feature assessment and scoring, about rules for feature weighting, and about the relationship between weighted features and overall effectiveness ratings. The experiment also provides data for comparing the subjects' curves with the curves extracted from our expert[1] and used to develop the training and test pictures. In addition, this experiment assesses the stability of the schema generated through the training procedure.

Methods

Materials. The materials for this experiment consisted of a set of 12 training pictures, 10 test pictures, a set of feature evaluation sheets, and the Raven Progressive Matrices Test (1958), which was used as a distractor task.

[1]A retired Navy commander

The training and test pictures illustrated threats capable of mounting all-out attacks of different effectiveness. These pictures were developed using a schema-like model of attack effectiveness. This model represented an expert's schema, and was developed by working with this expert. In developing this model, a set of pictures of all-out attacks and a set of features thought to be the basis of the effectiveness ratings for the attacks were developed. The expert was then asked to rate 1) the overall effectiveness of the attack shown in each picture, and 2) how characteristic each feature in the attack is of a very effective attack. The feature criteria curves that the expert was using were determined from these ratings; that is, the relationships between the physical properties of the picture (e.g., number of ships, submarines) and the subjective ratings for those features (e.g., many ships, submarines) was plotted. These curves were used to generate a new set of pictures. For each of these new pictures the overall assessment expected from the expert was predicted by converting the physical properties of the picture into subjective feature ratings and combining them using the geometric mean rule. The expert was again asked to rate the overall effectiveness of the picture. When discrepancies occurred between the predicted and actual ratings, the expert was queried for the cause of the discrepancies. The set of features and feature criteria curves were then modified based on that feedback. The entire process was repeated until the overall assessments given by the expert were predicted accurately from a weighted combination of the feature scores as calculated from our previously extracted feature criteria curves. It is important to note here that the wholistic preferences of the expert were never subject to question -- in each case the weights and criteria were changed, or identified, such that they matched these. Through this process three main features were finally identified as being important in determining the effectiveness of an all-out attack. These

were the overall ship strength, overall submarine strength, and the overall aircraft strength. For each of these features, the overall strength was a function of the number of platforms and of the number directions from which the platforms were able to attack. Any given picture depicts each of these features; each feature could be rated on a scale of one to ten in terms of how characteristic it is of an effective attack. The overall effectiveness of the attack results from a weighted geometric mean of the individual feature ratings. The feature criteria curves used to develop the materials for the low density attack (experiment 1) are shown in Table A-1 in Appendix 1.

The training and test pictures developed from this model illustrated the full range of possible attack effectiveness. Each picture was generated by choosing a level (high, medium, or low) for each feature and creating an attack representing these feature levels. The set of pictures was generated by varying the levels in a systematic way. The set of pictures included some where all the features were rated low, some where all the features were rated high, and some where the features represented a range from low to high. This process yielded a set of training and test pictures that had predicted overall effectiveness ratings throughout the full range, from one to ten. An example of a training picture, with its effectiveness rating and explanation for the rating, is shown in Figure 1. Table A-2 shows the design criteria for each test picture in the all-out attack experiment.

The feature evaluation sheets contained a list of the features relevant to determining the effectiveness of an all-out attack. For this feature, there were blanks to be completed regarding a) the extent to which each feature in the accompanying picture is characteristic of a very good attack (score 10), a very poor attack (score 1) or an intermediate attack (intermediate score); b)

how important that feature would be to an overall estimate of the effectiveness of the threat, and c) how confident the subjects were of the ratings they had just assigned for that feature. Each feature evaluation sheet was accompanied by one of the test pictures.

Procedure. The experiment began with a training session. In this session, subjects were provided with background material explaining the basic Battle Group scenario with which they would be working. Subjects were trained to recognize signs of all-out attacks by first instructing them on the signs of an impending attack and then by showing them six examples of all-out attacks. They were told how good each picture had been rated by an expert using a 10-point scale. They were also told which situation features were responsible for each example's rating. (See Figure 1 for an example.) Following this, they were shown an additional six pictures and asked to predict what each picture's rating would be. The actual expert's rating for each picture was then presented, along with an explanation of which situation features were responsible for the rating. After they had seen all twelve training pictures, they were asked to go back through all twelve pictures and predict the expert's rating until they could predict 9 of the 12 scores within one point on one pass through the pictures. Data from subjects who could not achieve this level of performance after three trials were not used in further analyses.

After the training session, subjects were shown ten test pictures. For each picture, they were asked to rate how effective the all-out attack pictured was and how confident they were that their effectiveness rating would match our expert's rating within one point. Each of these judgments was made on a 10-point scale.

When the ratings had been completed, subjects were asked to work on a series of puzzles, which were designed to serve as a distractor task. After working on these puzzles for twenty minutes, the subjects were given the feature rating sheets for the ten test pictures. For each feature in each picture, they were asked to rate a) to what extent the feature shown in the picture was characteristic of a very good attack; b) how important the feature would be to an overall assessment of the effectiveness of an all-out attack; and c) how confident they were of their ratings.

After all the feature sheets had been completed, the subjects were asked to make a new set of effectiveness and confidence ratings for the set of ten test pictures.

Subjects. The subjects were 25 undergraduate students at George Mason University in Fairfax, Virginia. Five of these subjects were unable to accurately predict the overall effectiveness ratings for the training pictures after three trials; their data were not analyzed further. The students received either course credit or payment for their participation in the study.

Results and Discussion

Our model assumes that schema are stable structures which are easily developed abstractions of examples. The data support this hypothesis. The mean number of trials required to reach criterion on the practice materials was 1.72, which suggests that the schema are easily learned, although there were five participants in this experiment who did not reach criterion by the third trial.

The stability of the schema can be seen in the stability of the overall effectiveness ratings, which were made independently, 60 minutes apart. These

ratings averaged over subjects can be seen in Table 1. The test-retest correlation for these ratings is extremely high ($\underline{r}$ (8) = 0.986, $\underline{p} < .01$), suggesting that the average rating for each picture is consistent over the 60-minute time span between the initial and final ratings. The stability of the ratings within individuals is also remarkable. Table 2 shows the frequency distribution of the difference between the first and second estimates of overall effectiveness for the first two experiments. Roughly one-third of the responses were identical on the two trials, 71% of the ratings on the second trial were within plus or minus one of the original rating; and 88% of the ratings on the second trial were within two points of the original response.

Another aspect of the data concerned the development of feature criteria curves for feature assessment and scoring. These curves relate the physical features of the picture (e.g., the number of ships) to the subjective ratings for that feature. Figure 5 shows the observed relationships between the actual number of platforms (ships, aircraft, and submarines) and the features "many ships/aircraft/submarines." Also shown are the curves used in the model to generate the experiment materials (labeled "target rating"). The curves monotonically increasing, (except for one point), with a consistent underestimation of the number of platforms relative to the curve used in the model.

The model also predicts that subjects' overall attack effectiveness ratings are approximated by the weighted geometric mean of the individual feature scores. We proposed that the weight assigned to each feature will be related to that feature's importance rating and that the geometric mean of the features so weighted will predict overall attack effectiveness more accurately than will the geometric means weighted in other ways. If these properties are true, then a) the correlation between the overall assessments and the geometric

PICTURE	LOW ATTACK CONDITION FIRST EVALUATION	LOW ATTACK CONDITION SECOND EVALUATION	HIGH ATTACK CONDITION FIRST EVALUATION	HIGH ATTACK CONDITION SECOND EVALUATION
1	9.7	9.6	9.9	9.9
2	7.9	7.5	8.2	7.2
3	6.7	6.8	6.25	6.5
4	8.1	8.35	7.85	8.1
5	6.25	6.05	5.5	6.4
6	5.2	5.4	4.85	5.55
7	4.6	5.1	4.1	5.4
8	5.4	5.85	4.7	5.45
9	3.15	3.8	2.15	3.3
10	1.65	2.45	1.65	2.26
Average	5.86	6.09	5.56	6.05

TABLE 1. Average all-out attack effectiveness ratings for first and second evaluations of all-out attacks, low density case Expt 1 and high density case Expt 2.

Difference between first and second attack estimates	# Responses	Frec. Responses
-5 or 6	4	.01
-4	2	.005
-3	8	.02
-2	26	.06
-1	63	.16
0	130	.32
1	94	.23
2	44	.11
3	19	.05
4	10	.025
5	4	.01

TABLE 2. Frequency distribution for the difference in individual scores between the first and second "all out attack" effectiveness ratings. There are 400 responses from 40 subjects, each judging 10 test pictures.

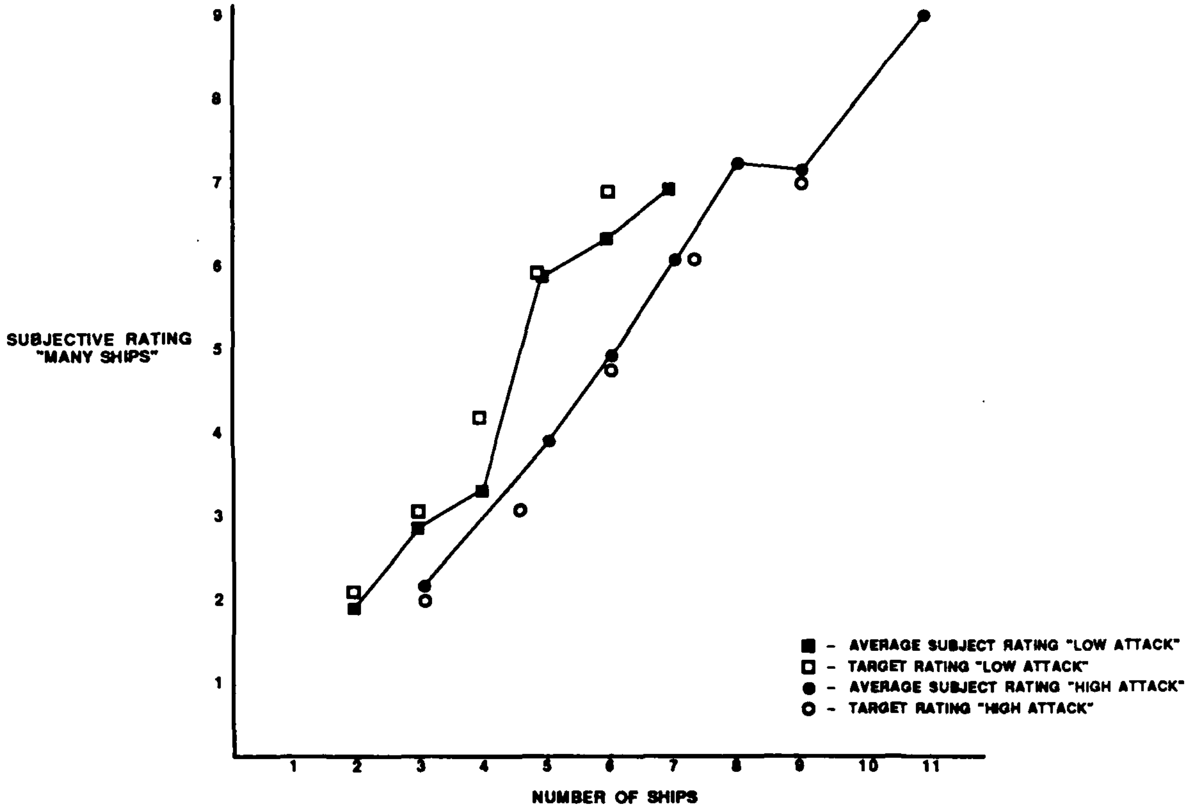


Figure 5a. Feature Scaling Curves Relating Subjects' Ratings "Many Ships" To Number Of Ships In Test Pictures.

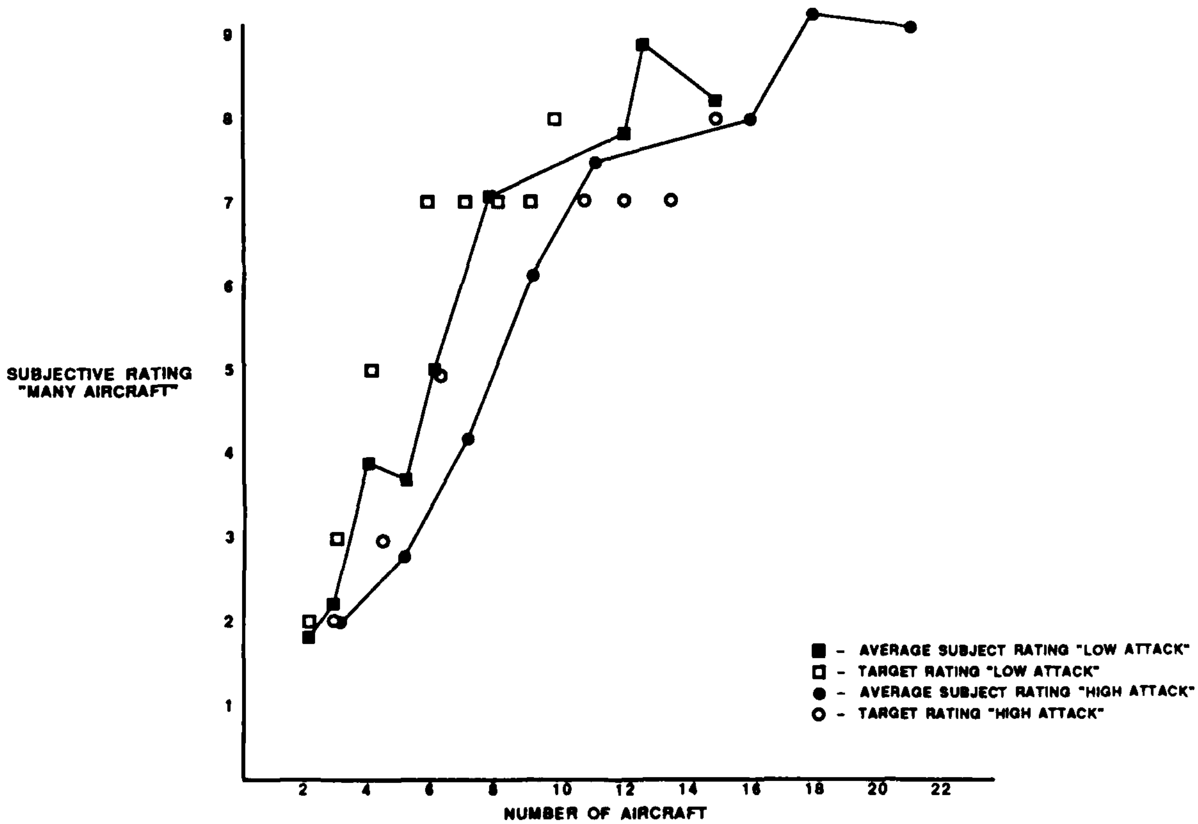


Figure 5b. Feature Scaling Curves Relating Subjects' Ratings "Many Aircraft" To Number Of Aircraft In Test Pictures.

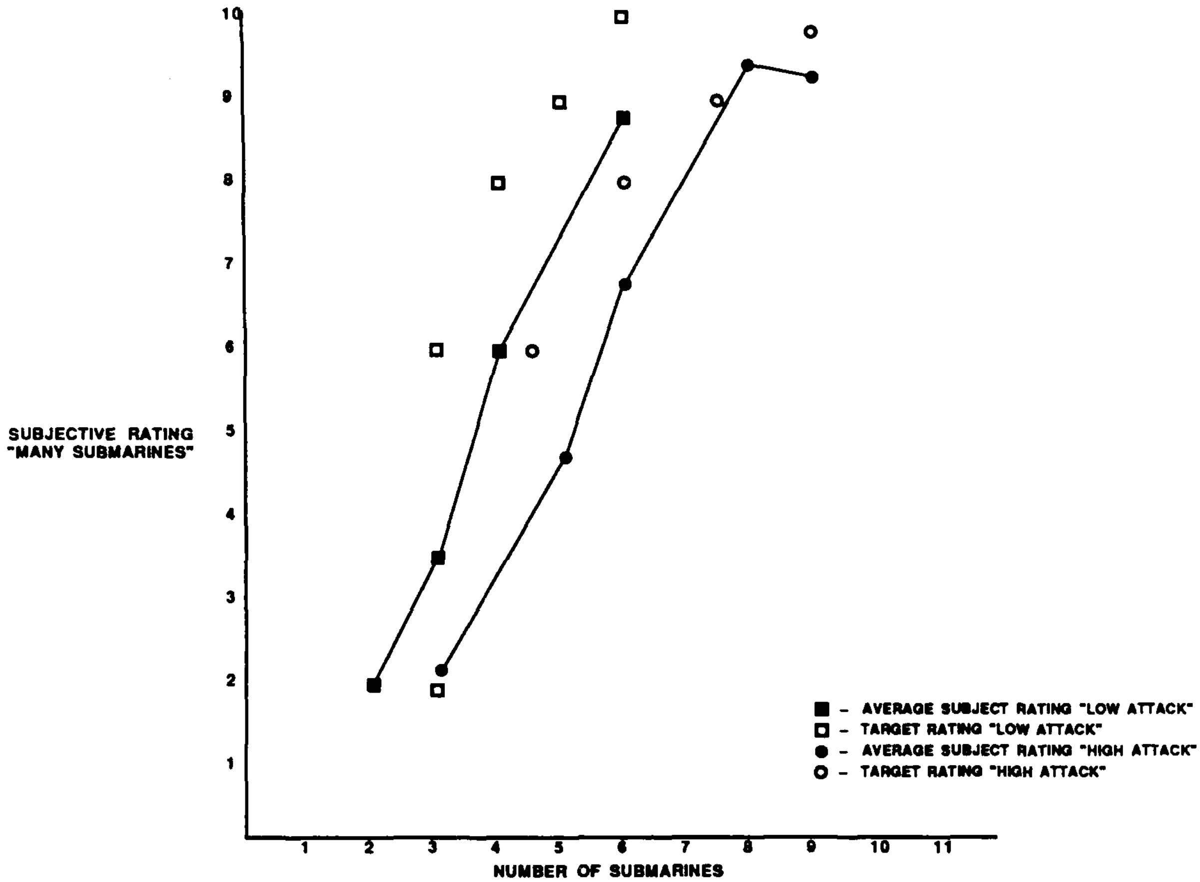


Figure 5c. Feature Scaling Curves Relating Subjects Ratings "Many Subs" To Number Of Subs In Test Pictures.

mean of the individual feature ratings should account for a significant proportion of the variance in the overall ratings; and b) this relationship should be improved by weighting the individual features by their importance rating before calculating the geometric mean.

In this experiment, subjects rated the features, their importance, and the overall effectiveness of the pictures independently. They rated overall effectiveness at a different time from their other ratings, and did not have available a record of their ratings made at different times. We calculated unweighted and weighted geometric means of the individual feature ratings. The weights were attained from the importance ratings by converting each rating of high, medium, and low to 3, 2, and 1, respectively, and then squaring this number (a feature rated high counts nine times as much as one rated low). The unweighted and weighted geometric means accounted for 92 and 97 percent of the variance in the overall assessments ($\underline{r}$ (8) = 0.960 and 0.983 $\underline{p} < .01$). The relationship between the weighted features and the overall assessments for experiments 1 and 2 can be seen graphically in Figure 6. Table 3 presents the correlations and regression lines for each individual subject in experiments 1 and 2. This table shows that the relationship shown in Figure 6 for data averaged over subjects is also observed for individuals.

While the weighting did not increase by much the already high variance accounted for by the unweighted geometric means of features, weighting did significantly reduce the absolute difference between the overall effectiveness ratings and the feature geometric means. Table 4 shows the average overall effectiveness rating for each picture and the predicted overall effectiveness rating using: a) the unweighted geometric mean, b) the geometric mean weighted by the squares of the importance ratings, and c) the geometric mean weighted by

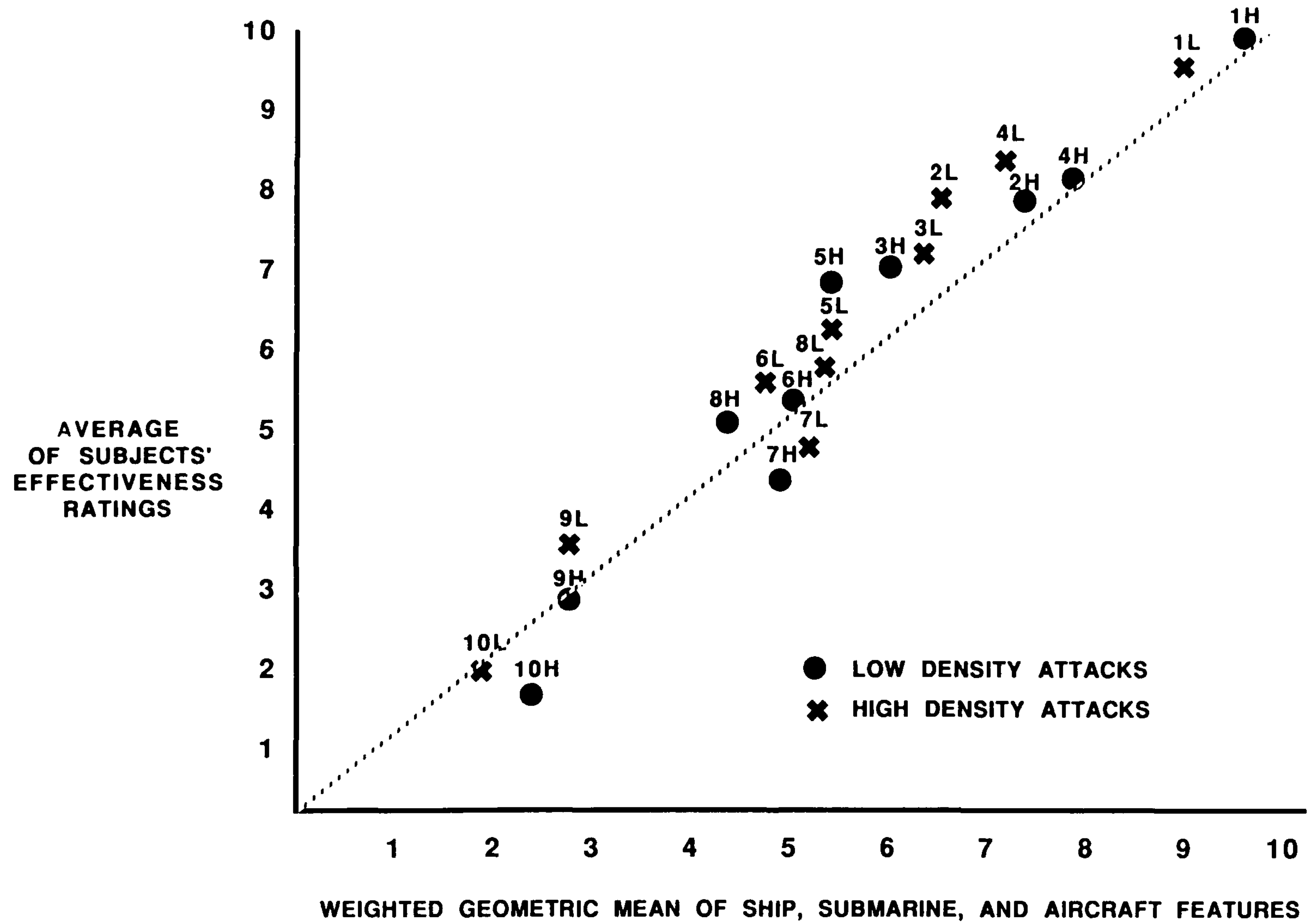


Figure 6. Average Effectiveness As A Function Of Weighted Geometric Mean Of Feature Ratings Numbering Shows Picture Number, Experiment Condition. Geometric Mean Is Weighted By Square Of Importance Ratings

ONR-003

	LOW ATTACK CONDITION			HIGH ATTACK CONDITION			
SUBJECT	WEIGHTED CORRELATION	LEAST SQUARES COEFFICIENTS a	b	SUBJECT	WEIGHTED CORRELATION	LEAST SQUARES COEFFICIENTS a	b
1	0.89	1.11	0.29	1	0.90	0.95	1.03
2	0.80	0.92	0.56	2	0.93	0.75	1.55
3	0.82	1.04	1.33	3	0.63	0.71	1.77
4	0.88	0.95	0.12	4	0.96	0.99	0.59
5	0.77	0.92	0.08	5	0.81	0.63	1.68
6	0.83	0.69	1.97	6	0.87	1.04	1.00
7	0.9	1.21	0.40	7	0.89	0.94	1.62
8	0.92	1.06	-0.11	8	0.95	1.09	-1.32
9	0.87	0.92	1.16	9	0.78	0.84	0.61
10	0.82	0.88	2.28	10	0.86	0.81	0.71
11	0.82	0.62	3.43	11	0.92	0.87	0.92
12	0.87	0.98	0.51	12	0.87	0.99	0.02
13	0.90	0.99	1.26	13	0.91	0.92	0.68
14	0.84	0.77	0.29	14	0.65	0.67	2.75
15	0.88	0.62	2.25	15	0.92	1.17	1.26
16	0.81	0.88	0.26	16	0.83	0.68	2.36
17	0.46	0.54	2.68	17	0.80	0.86	0.44
18	0.85	1.01	2.19	18	0.93	0.84	0.91
19	0.81	0.69	2.13	19	0.91	0.84	1.10
20	0.95	0.93	0.92	20	0.94	1.13	-1.50
Average	0.83	0.89	1.20	Average	0.86	0.89	0.96
S.D.	0.10	0.18	1.03	S.D.	0.09	0.15	0.92

TABLE 3. Individual correlation coefficients and least squares coefficients on all-out attack. Correlation is between assessment of attack effectiveness and geometric means weighted by squares of importance rating. Least squares fits line:
Average attack effectiveness = a x geometric mean of features + b.
Subject pools for "low" and high cases are different.

EXPERIMENT 1: LOW DENSITY

PICTURE	Avg Eff	Unwgted	Wgted by Sq of Importance	Wgted by Feature Rating	Avg. Eff.- Unwgted	Avg. Eff.- Wgted by Sq.	Avg. Eff- Wgted By Rating
1	9.65	8.93	9.1	9.06	.72	.55	.59
2	7.70	6.27	6.44	6.43	1.43	1.36	1.37
3*	6.75	6.03	6.33	6.36	.72	.42	.39
4*	8.23	6.06	7.45	7.22	2.17	.78	1.01
5*	6.15	3.97	5.26	5.57	2.18	.89	.58
6	5.30	3.85	4.42	4.37	1.45	.88	.93
7*	4.85	4.02	5.13	5.37	.83	-.28	-.52
8	5.62	4.07	5.04	5.10	1.55	.58	.52
9	3.47	2.46	2.83	3.2	1.01	.64	.27
10	2.05	1.79	1.94	1.98	.26	.11	.07
Avg.	5.97	4.74	5.39	5.46	1.23	.56	.51

EXPERIMENT 2: HIGH DENSITY

1	9.9	9.44	9.56	9.5	5.6	.34	.4
2	7.7	6.91	6.98	7.19	.79	.72	.51
3*	6.38	5.62	5.74	5.95	.76	.64	.43
4*	7.97	6.94	7.65	7.82	1.03	.32	.15
5*	5.95	4.57	5.30	6.29	1.38	.65	-.34
6	5.2	4.85	4.95	5.42	.35	.25	-.22
7*	4.75	4.26	4.82	5.90	.50	-.07	-.85
8	5.02	4.11	4.47	5.33	.91	.45	-.31
9	2.97	2.64	2.74	3.22	.33	.23	-.25
10	1.95	2.15	2.20	2.38	-.20	-25	-.23
Avg.	5.74	5.15	5.45	5.9	.59	.29	-.16

TABLE 4. Geometric means of all-out attack feature scores, unweighted, weighted by square of importance, and weighted by feature rating. *Pictures with ship, submarines, and aircraft that differ most in strength.

the actual feature assessment ratings. The remaining columns show the difference between the overall effectiveness ratings and the ratings predicted by each weighting process. It can be seen that using a mean weighted by importance reduced the average error from 1.23 rating points (on 10-point scale) to 0.56 rating points, a 54% reduction in average error.

Another interesting result is the similarity between the means weighted by importance versus the means weighted by the actual feature score. These two weighting procedures give virtually the same predicted overall rating. This result suggests that importance ratings for these features are derived from the assessed effectiveness of the feature; that is, features rated more characteristic of effective attacks were rated more important than were features regarded as less characteristic of effective attacks. Because in this experiment rated importance seems related to feature weight, this result also suggests that the criteria curves used to generate feature scores were also used to generate feature weights.

The data in this experiment indicates that the subjects' schema are reasonably accurate. These data show that subjects' estimates of overall attack effectiveness approximate the ratings predicted for the test pictures based on our model of the expert's knowledge. Table 5 compares for each picture the target ratings calculated from the model with the observed attack rating averaged over subjects and trials for each experiment for each picture. The correlation between the targets and observed ratings was significant ($\underline{r}$ (8) = 0.942, $\underline{p} < .01$).

PICTURE	ALL-OUT ATTACKS			BARRIERS	
	MODEL	ACTUAL LOW	ACTUAL HIGH	MODEL	ACTUAL
1	9	9.65	9.90	3	4.15
2	7.95	7.7	7.7	5.21	4.9
3	7.25	6.75	6.38	5.6	5.55
4	6.8	8.23	7.97	8.4	8.65
5	5.85	6.15	5.95	3.56	4.25
6	4.95	5.30	5.20	1.7	2.30*
7	4	4.85	4.75	7.6	7.85*
8	4	5.62	5.02	10	9.70*
9	3.1	3.47	2.97	1.8	2.9*
10	2	2.05	1.95	6.8	8.10*
Average	5.5	5.97	5.74	5.36	5.83

* also in training set

TABLE 5. Comparison of subjects' Effectiveness Ratings with ratings produced by model of the expert's knowledge. "Actual low" and "actual high" refer to the low density and high density attack experimental conditions.

Experiment 2: All-Out Attack, High Density

This experiment was designed to provide data on the feature criteria curves used for feature assessment. The model proposes that such curves will be inferred from the examples provided during training. To test this proposition, each training and test picture in experiment 1 was modified to contain 50% more ships, aircraft and submarines, and then used in this experiment. If the feature assessment curves are derived solely from the examples provided in training, then the feature scaling curves derived from experiments 1 and 2 should be the same except for an x-axis scaling factor of 50%.

In addition, since the procedure was identical to that used in experiment 1, the experiment provides a replication of the data collected on feature scaling, weighting and combination rules, on the extent to which subjects learn the model we used to develop the pictures, and on the stability of these schema.

Method

Materials. The materials for this experiment were the same as those used in experiment 1 except that each of the training and test pictures in experiment 2 contained 50% more platforms (ships, submarines, aircraft) than the corresponding picture in experiment 1. The additional platforms were placed close to the original platforms in order to minimize any effect on the perceived number of attack axes.

Procedure. The procedure for this experiment was identical to that described for experiment 1.

Subjects. The subjects were 20 undergraduate students at George Mason University in Fairfax, Virginia. The students received either course credit or payment for their participation in the study.

Results and Discussion

This experiment was designed to work with experiment 1 in order to test the properties of the feature criteria curves used for feature assessment. The experiment also provided data that independently supports many of the hypotheses examined in the first experiment.

In this experiment, all subjects reached criterion. The mean number of trials required to reach criterion was 1.50. The schema stability results of this experiment resemble those of the first. The average effectiveness ratings for each picture on each trial are shown in Table 1. The test/retest correlation between these ratings is significant (r (8) = 0.977, p < .01), and indicates that the ratings are stable over time.

The data pertaining to feature weights and the relationship between the assessed attack effectiveness and feature geometric means also resembled those from the first experiment. We again found that both the unweighted and weighted geometric means accounted for more than 90% of the variance in the overall effectiveness ratings and that the weighted geometric means accounted for somewhat more variance than did the unweighted. The geometric mean of the weighted individual feature rating is plotted against the overall effectiveness ratings in Figure 6. Again, the importance of weighting is shown by the data in Table 4, as discussed in the results section for experiment 1.

Table 5 compares the target ratings calculated for each picture from the model with the observed attack effectiveness ratings averaged over subjects and trials for each experiment. Again, the correlation between the targets and observed ratings for this experiment was significant (r (8) = 0.949 p < .01).

Data collected in this experiment reveal how the training pictures interact with "commonsense" knowledge to affect the feature criteria curves used for feature assessment. Our initial hypothesis was that feature criteria curves are abstracted solely from the examples given. If subjects' criteria curves derive only from the examples given in training, then the overall assessments for the feature "many platforms" should be the same for corresponding pictures in experiments 1 and 2 even though the number of platforms has changed. If this were the case, then the criteria curves inferred from the pictures used in experiment 2 (with 50 percent more platforms) would differ from the curves inferred from the pictures used in experiment 1 only by a 50% scaling factor on the x-axis. The results suggest that this was not the case. The curves from experiment 2 (Figure 5) reflected only part of the 50% increase expected.

Although the feature criteria curves depend only in part on the training pictures, the overall attack assessments seem to depend solely on the training. The attack evaluation scores for corresponding pictures in experiments 1 and 2 are virtually the same.

Experiment 3: Barriers

This experiment served several functions. First, it reexamined several issues in experiments 1 and 2, providing a second example of feature criteria curves, feature weighting rules, and feature combination for situation and assessment. Second, it examined several new issues, providing data to determine whether people use functional as well as physical properties of objects to make their judgments, and to determine whether people would apply their everyday knowledge of objects to existing schema. In addition, it was designed to examine the relationship between similarity assessments of pairs of situations and their constituent feature ratings.

Methods

Materials. The materials for this experiment consisted of a set of ten training pictures, seventeen test pictures, the feature evaluation sheets, feature comparison sheets, and the Raven Progressive Matrices Test (1958), which was used as a distractor task.

The set of training pictures and one set of test pictures were constructed from a model of barrier goodness that specified feature criteria curves and a feature weighting rule. Unlike experiment 1, this model was not developed from an expert's model for barrier evaluation. This model specifies two features relevant for barrier effectiveness assessment, the length of the barrier and the solidity of the weakest part of the barrier. Two feature criteria curves relate the measurable properties of the picture's features (distance between two end ships/subs, distance between two platforms on either side of the largest internal gap) to subjective feature scores (barrier length and solidity). These relationships are shown in Table A-3. Overall barrier effectiveness is calculated from the weighted geometric mean of feature scores attained from these criteria curves. Since in our model of barrier effectiveness a barrier was only as strong as its weakest link, the weaker feature is weighted by .75 and the stronger by .25.

For the first part of this experiment, 15 pictures were developed; five pictures were shown during training only, five were shown during test only, and five were shown both as training and test pictures. The overall ratings ranged from two to ten for both the training and test pictures. An example of a test picture is shown in Figure 7. Table A-4 shows the overall design for the barrier pictures.

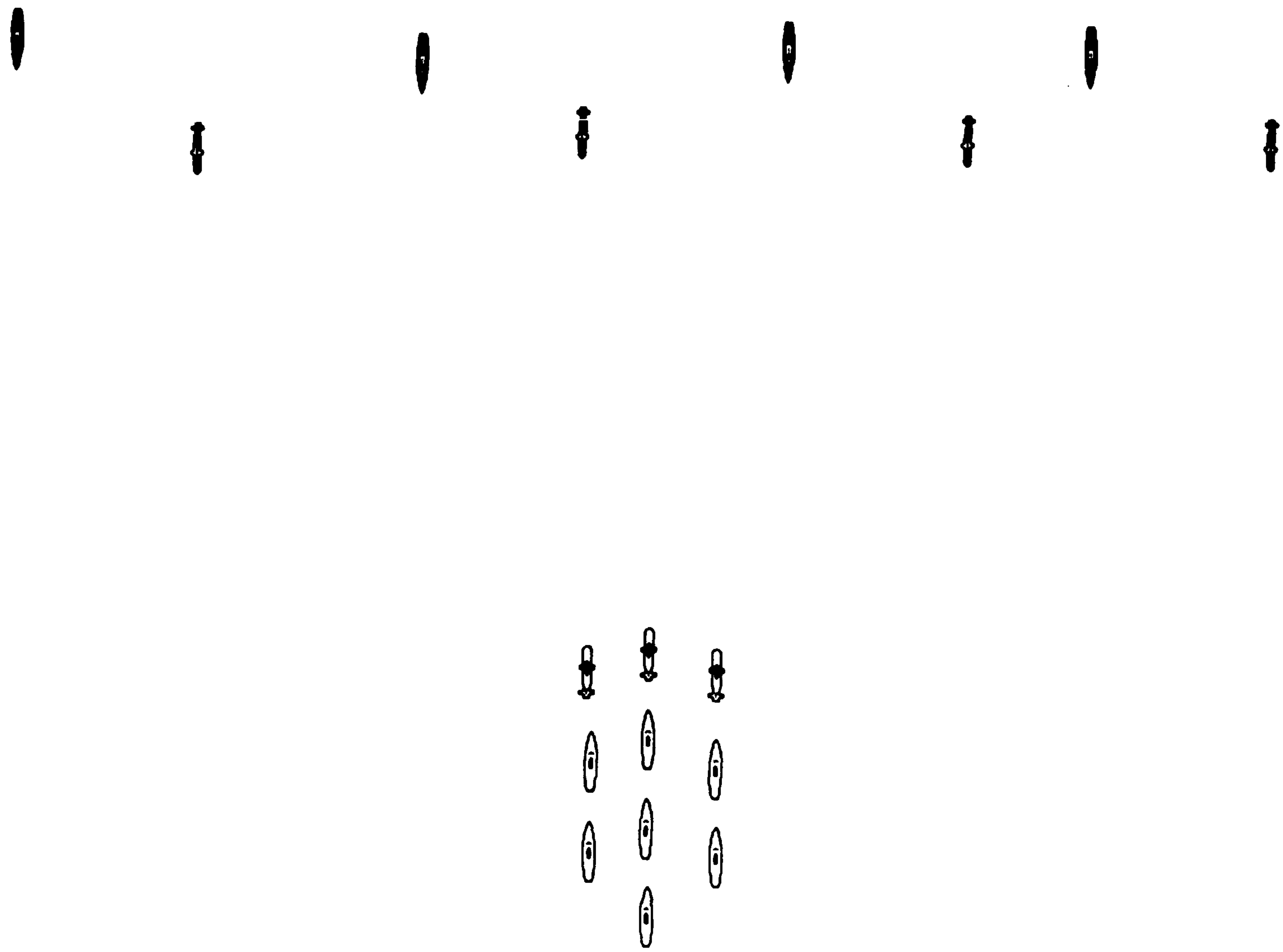


Figure 7. An example of a training picture for experiment 3, Barriers.
During training subjects were told:
Picture 1: Effectiveness = 10. The barrier is both long and solid. The ships at the two ends are sufficiently far apart to make the barrier difficult to go around. The platforms are close enough together throughout its entire length to make passage through the barrier very difficult.

Seven pictures were developed for the second part of the test. These pictures were modifications of pictures shown in the first part. Five of the pictures were modified by adding either an island or peninsulae to the picture. This procedure created pictures which physically matched one of the original test pictures in terms of number and location of platforms, but functionally matched a second original test picture, in terms of length and solidity of the barrier. An example of a test picture from this set and the physically and functionally equivalent pictures is shown in Figure 8. Two other new test pictures were created by taking two of the original test pictures and moving the platforms to one side, so that they were no longer centered in front of the battle group. Again, this created pictures which were physically similar to the one of the original test pictures, but functionally similar to another original test picture.

The feature evaluation sheets for this experiment were similar to those used for the all-out attacks. This sheet listed six features intended as "physical", "intermediate" and "functional" representations of the barrier length and solidness. The functional features are "barrier is hard to go around" and "barrier is hard to go through". The intermediate features are "barrier length" and "barrier solidness". The physical features are "distance between end ships/subs" and "distance between ships/subs on either side of the largest internal gap".

Feature comparison sheets were used to assess the similarity between features depicted in pairs of test pictures. Each page contained two pictures of barriers, one from the original test set and one from the second test set. Below the pictures was a list of the six features relevant to determining the

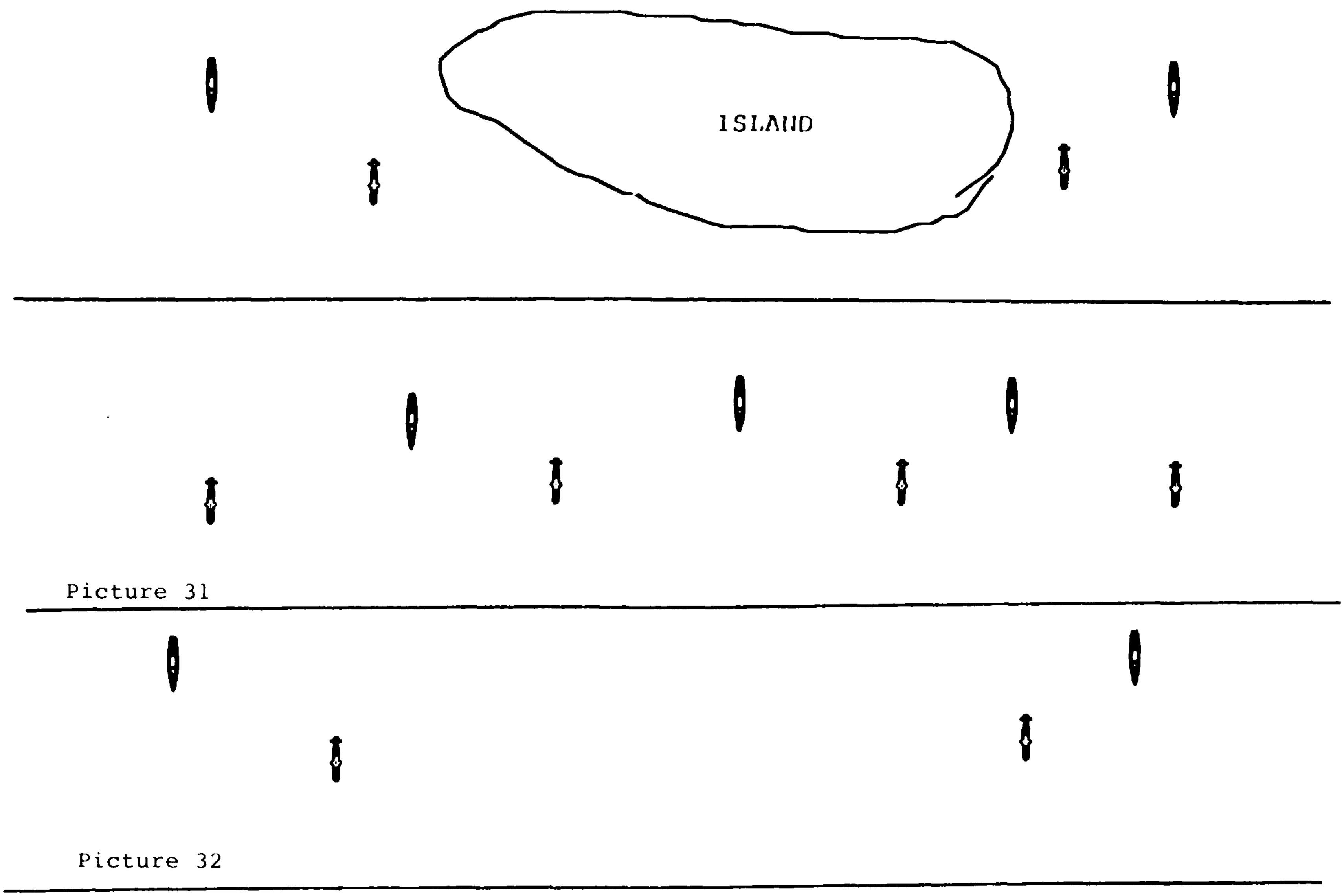


Figure 8. An example of a barrier test picture with an island. Picture 31 is functionally equivelent to the barrier. Picture 32 is its physical counterpart if the island is disregarded.

effectiveness of a barrier. Subjects were asked to rate how similar the first barrier was to the second with respect to each of these features.

Procedure. The procedure for this experiment was the same as that described for Experiment 1 with the following exceptions. Subjects in this experiment saw only 10 training pictures and worked on the puzzles which served as a distractor for only 10 minutes. After completing the feature ratings for the test pictures, the subjects were asked to make effectiveness and confidence ratings on seventeen additional pictures, and later were asked to complete feature evaluation sheets on each of the new test pictures. Finally, they were presented with pairs of barrier pictures and were asked to rate the similarity of the each of the features presented in the two barriers.

Subjects. The subjects were 20 undergraduate students at George Mason University in Fairfax, Virginia. The students received either course credit or payment for their participation in the study.

Results and Discussion

Again in this experiment, all of the subjects reached criterion. The mean number of trials require to reach criterion was 1.15. In addition, the data from this experiment suggest that subjects abstracted from the examples the model used to develop the pictures. The geometric mean of the features' scores, weighted according to the weighting rule used to construct the pictures, accounted for 88 percent of the variance in the overall effectiveness ratings ($r(8) = 0.937$, $p < .01$). Figures 9 and 10 depict this relationship for each of the two test sets. The subjects' assessments of barrier quality for test pictues seen earliwe in training and the new test pictures not seen in training both approximated the model's assessments equally well. This result suggests

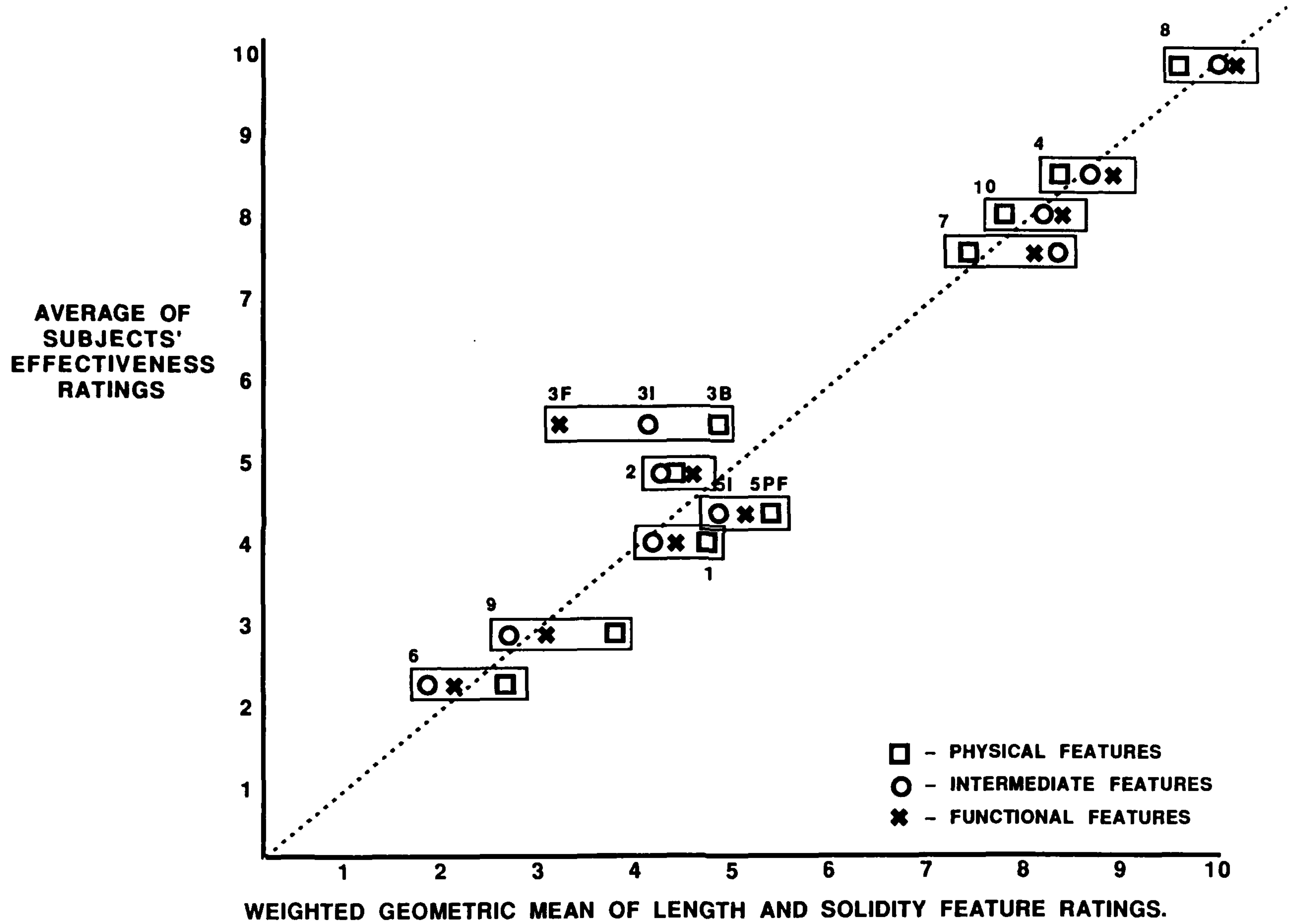


Figure 9. Average Assessments Of Barrier Effectiveness As A Function Of Weighted Geometric Mean Of Physical, Intermediate And Functional Features. Test Pictures With Centered Ships/Subs
Weight: .75 For Weaker Feature .25 For Stronger Feature.

ONR 005

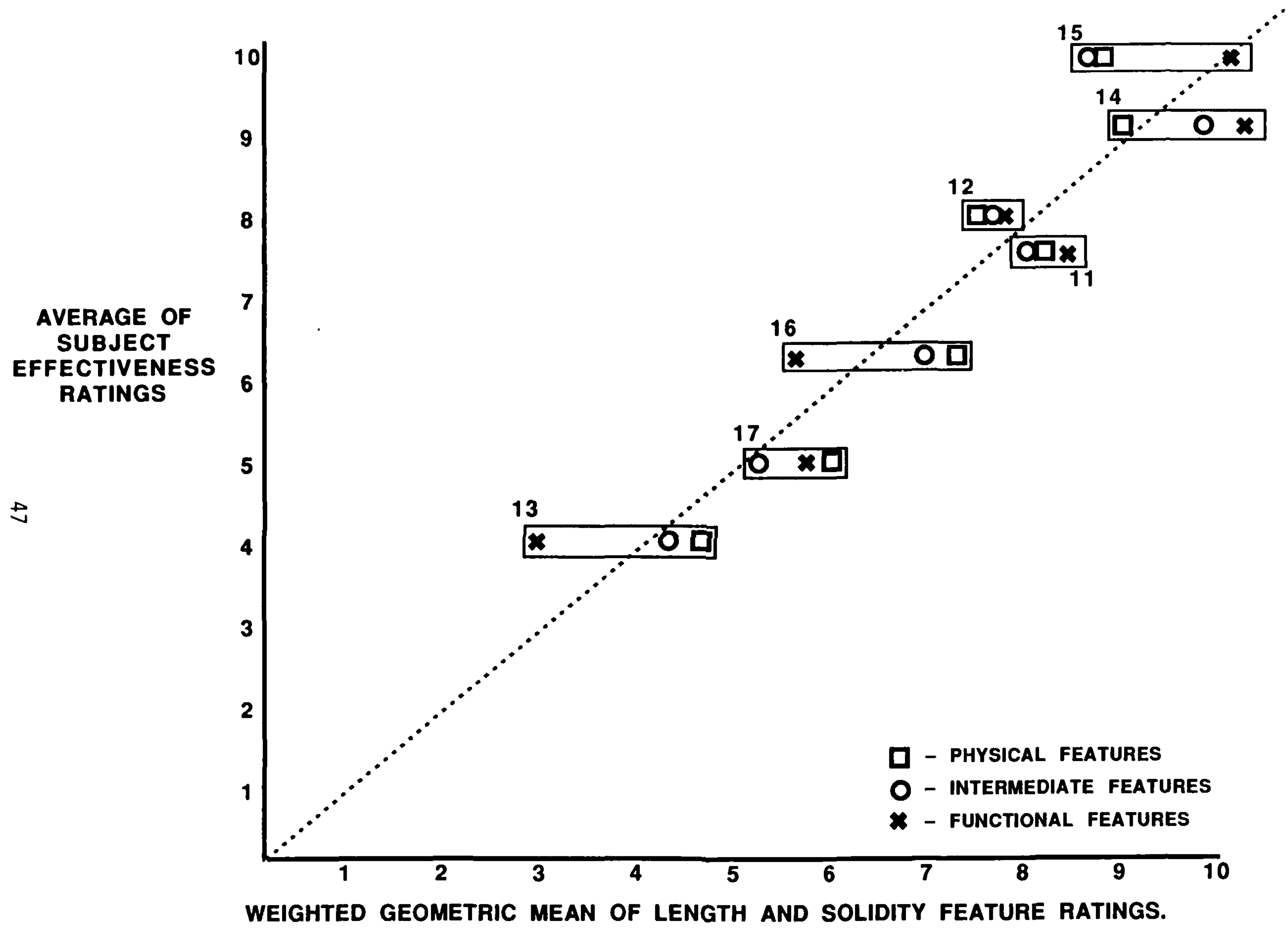


Figure 10. Average Assessments Of Barrier Effectiveness As A Function Of Physical, Intermediate And Functional Features. Test Pictures 11-15 Islands And Peninsulae. Pictures 16 And 17, Off-centers

ONR-006

that subjects have internalized a barrier assessment process and are not just remembering the pictures that they saw during the training period.

The data on functional substitution indicate that newly formed schema interact closely with other knowledge in memory. During the training for this experiment, subjects never saw islands or peninsulae as part of the barrier. The data from the ratings, however, suggest that subjects incorporated their existing knowledge of the properties of land masses into their overall effectiveness ratings.

Table 6 shows a schematic representation of each test picture, and the effectiveness ratings given to that picture and to its physical and functional equivalents. It can be seen from the table that in the island/peninsula group the functionally equivalent pictures provided a better match to the initial ratings than did the physically equivalent pictures, except for one picture. A closer examination of this picture suggests that the functional equivalent chosen for this picture was not appropriate because the new test item allows safe passage through the internal gap, while such safe passage is not provided by the proposed functional equivalent for this picture.

The "off-center" test pictures, where the ships were displaced to the side, did not produce as clear results on this functional/physical equivalency. For those pictures the barrier ratings do not match the ratings of the functional equivalents better than do the ratings of the "look alikes".

A comparison of the subjects' responses for physical, intermediate, and functional features suggests the point in the information processing sequence at which subjects use the functional properties of land masses in their barrier assessments. Table 7 shows that the weighted geometric means of the physical,

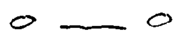
PICTURE	DESCRIPTION	RATING	RATING OF PICTURE IT "LOOKS LIKE"	RATING OF PICTURE IT "FUNCTIONS AS"
11		7	4.9	8.1
12		7.1	4.15	8.1
13		3.75	2.3	7.85
14		8.25	5.6	8.65
15		8.7	2.9	9.7
16		6.25	8.1	4.15
17		4.8	4.9	2.9

TABLE 6. Comparison of ratings of off-center, island, and peninsula barriers with their functional and "looks like" counterparts. Pictures 11 through 15 have islands and peninsulae added. In pictures 16 and 17 the barriers are displaced off-center.

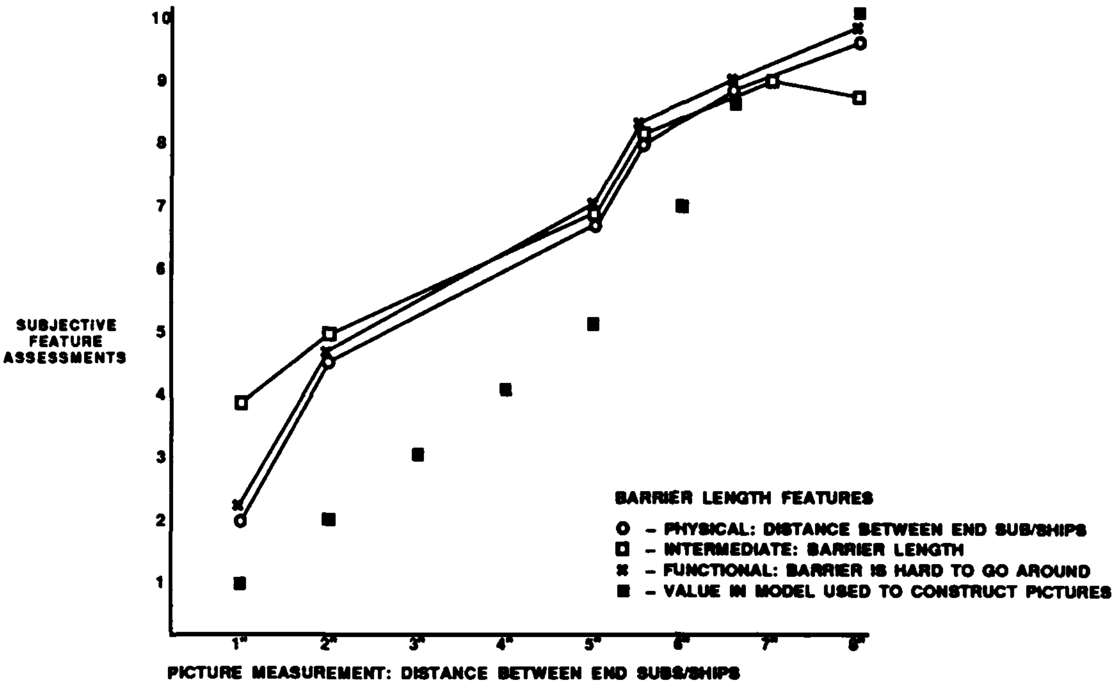
		AVERAGE PICTURE ASSESSMENT	WEIGHTED GEOMETRIC MEANS OF FEATURES: PHYSICAL	INTERMEDIATE	FUNCTIONAL
CENTERED BARRIERS WITH ONLY SUBMARINES AND SHIPS	AVERAGE ASSESSMENT	5.83	5.82	5.74	5.77
	DEVIATION FROM ASSESSMENT		.49	.66	.78
OFF CENTER BARRIERS, BARRIERS WITH ISLANDS AND PENINSULAE	AVERAGE ASSESSMENT	6.55	6.68	6.67	6.62
	DEVIATION FROM ASSESSMENT		.59	.48	.61

TABLE 7. Physical, Intermediate, and functional features as predictors of subject's assessments of barrier effectiveness. Average is average over test pictures. Geometric mean weights are .75 for weaker feature and .25 for stronger feature.

intermediate, and functional features predict the overall effectiveness ratings about equally well. Because the physical features predicted overall assessments as well as did the functional assessments, subjects must have taken the islands and peninsulae into account when rating the distances between the two end ships/subs or the two platforms on either side of the largest internal gap. This result implies that the conversion of land masses to ship equivalents occurs before the criteria curves were applied to the measureable feature properties.

The actual feature criteria curves used by the subjects are shown in Figure 11. The figure shows that the subjects' curves do not replicate curves used in the model. The curves for barrier length are generally too high, while the curves for barrier solidness are generally too low. This lack of agreement with the model is not surprising. During the training session, the subjects are not told how much each feature contributes to the overall effectiveness rating. Since there are many combinations of length and solidness ratings whose geometric mean aproximates the overall effectiveness ratings, it is not possible for the subjects to infer the combination used by the model.

As in experments 1 and 2, the weighted geometric means accounted for most of the variance in the barrier effectiveness assessments. Here they accounted for 93 percent of the variance ($\underline{r}$ (8) = 0.965 $\underline{p}$ < .01) in the target ratings. Again, the weighted means more accurately predict the overall effectiveness ratings than the unweighted means. As shown in Table 8, the unweighted mean overestimates the effectiveness ratings by 0.44 for the standard test pictures and 0.6 for the island/off-center test pictures. Using the weighting procedure reduces the error to -0.09 for the standard test pictures and to 0.12 for the island/off-center test pictures.



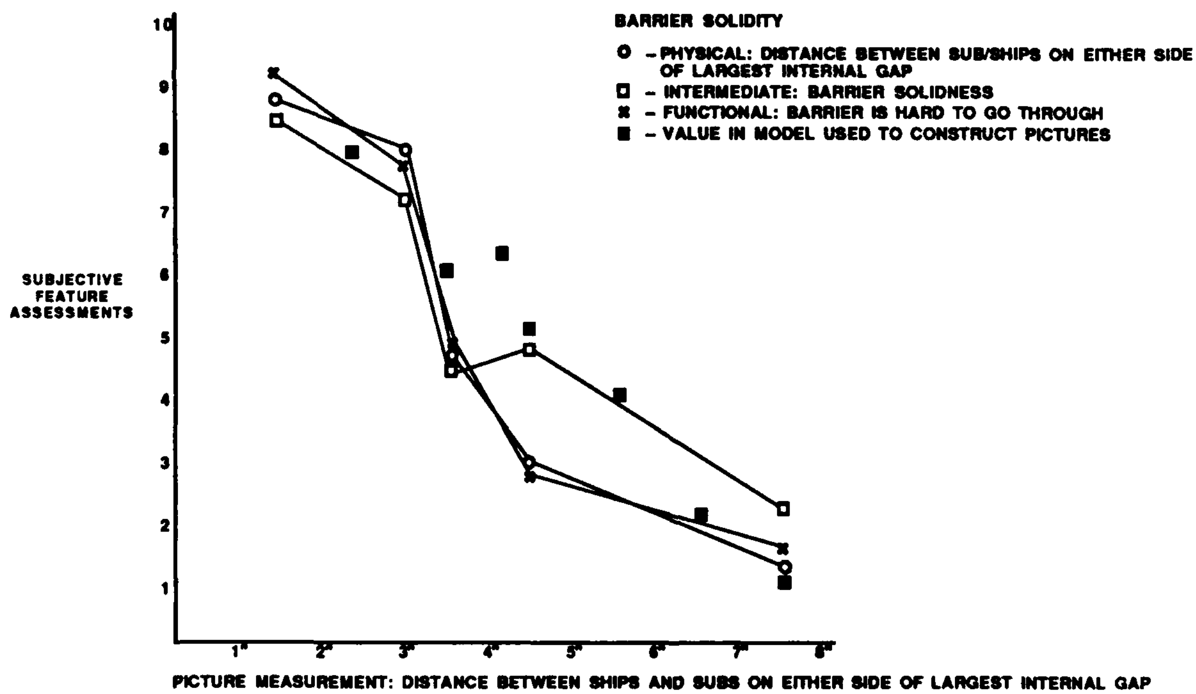


Figure 11. Feature Criteria Curves Relating Physical, Intermediate And Functional Feature Assessments To Measurable Properties Of The Barrier Features.

ONR-884

FEATURE: LENGTH AND SOLIDITY

	PICTURE RATING	UNWEIGHTED GEOMETRIC MEAN	WEIGHTED GEOMETRIC MEANS By IMPORTANCE	By STRENGTH	By 75-25 rule
PICTURES 1-10 STANDARD	5.83	6.27	6.45	7.3	5.74
PICTURES 11-17 OFF CENTERS/ ISLANDS/ PENINSULAE	6.55	7.15	7.23	7.64	6.67

TABLE 8. Average ratings compared with unweighted and weighted geometric means of feature ratings.

Table 8 also shows that in the barrier pictures importance ratings cannot be equated with the feature strength as measured by feature assessment scores. Using importance as a weighting factor in this experiment led to a worse fit between feature geometric mean and barrier assessment than using the unweighted mean. This result suggests that, in general, importance rating is a combination of the true feature weight and feature strength. In this case, the weighting rule used to construct the training and test pictures produced the best fit. This result shows that subjects inferred from the training that barrier quality is primarily determined by its weaker component.

The analysis of feature similarity data failed to reveal any special relationship between the ratings of features in test pictures and the similarity of these features to features in the training pictures. As expected, features rated similar received similar feature ratings. Those rated dissimilar, however, could also receive similar ratings. This result presumably reflects the fact the features rated equally strong can be strong in different ways.

General Discussion

The data in these experiments resolved most of the issues described under "critical model issues". This discussion reviews the data support for the alternatives described for each of these critical issues.

Ease of learning, stability, and accuracy of schema abstracted from examples. The data pertaining to these issues confirm the proposed schema model of situation assessment. In each of the three experiments subjects acquired the schema easily, as measured by trials to criteria during training. In each, they retained a stable schema throughout the experimental session, as measured by the consistency of situation assessments made at different times. In each, their schema captured the model used to develop the training pictures, as measured by the similarity of their assessments to the model's ratings. The inclusion of training pictures in the barrier test set provided additional evidence that subjects had based their assessments on a schema-like model rather than by remembering specific instances. The subjects' estimates of situation quality for the pictures not seen in training approximated the model's rating for those pictures as closely as did their estimates for the pictures seen in training.

Assessment of feature relevance and functional substitution. The model proposed that subjects, when shown pictures that contain familiar objects not in the training pictures, would consider these objects in their situation assessments. The model proposed that people would use general knowledge about islands and peninsulae (ships cannot pass over land) and off-centeredness (easier to go around) in evaluating the quality of barriers. The proposition that subjects' schema for barriers would accommodate islands and peninsulae and unusual barrier placement was confirmed. With one exception, people's

assessments of barrier quality was much closer to the barrier rating of the functional equivalent than of the look-alike. It is easily argued, however, that in the one case where their assessment was closer to the look-alike, the barrier would not in fact function like the presumed functional equivalent.

The data do not support the conjecture that people attempt to evaluate the barriers using physical features, and use functional features only when the physical features lead to an assessment of low barrier quality. This conjecture would have been supported had people's assessments of barrier quality been predicted from the functional features (hard to pass through, hard to go around) but not from the physical features (distance between ships on either side of largest internal gap, distance between end ships). Indeed, the data showed no indication that functional features alone contribute to barrier assessment when new objects are introduced into the barriers. Rather, physical and functional features always were equally good predictors of overall assessments. The means of physical and functional feature scores were extremely close for all pictures.

On the other hand, the data do support the existence of a very early information processing step in which new objects not seen in training are mentally replaced with a functional equivalent of objects seen in the training. The fact that physical and functional feature scores were so close suggests that the subjects, when answering questions about ship distances, were already taking into account the effects of islands and peninsulae, perhaps by treating the islands and peninsulae as additional ships. The early functional replacement of "nonstandard" objects with standard ones is attractive, for it seems to increase the general applicability of schema while minimizing the schema memory requirements. If nonstandard units are converted to the units used by the physical feature criteria curve, then the curve data can be stored more economically than

if a separate curve is required for every kind of object that can contribute to barrier length or solidness.

Feature assessment and scoring--use of criteria curves. Feature scoring is the conversion of measurable picture properties, such as the number of aircraft in an all-out attack, to a subjective estimate of the contribution of that feature to a strong all-out attack. The model proposes that feature scoring is accomplished by evaluating features by means of the criteria curves stored within the schema.

The data suggest that feature scoring is an important step in situation assessment. The curves themselves are simple monotonic functions of the measurable feature property, and the feature values obtained from these curves seem to be used in the overall assessments.

A comparision of the feature criteria curves attained in the first two experiments, the low and high density all-out attacks, shows that the feature criteria curves inferred by the subjects are derived from a combination of the training materials and general knowledge not part of the training.

The all-out high density and all-out low density experiments differed only in the numbers of platforms in the pictures. Every test and training picture in the high set was identical to a corresponding picture in the low set, except that the high set contained 50% more of each platform type. The words used to describe the pictures were identical, and the geographic arrangement of platforms were as similar as possible. In these two experiments the average of subjects' ratings of corresponding pictures in the two experiments were virtually identical, as were the weighted geometric means of the feature scores. Because the number of platforms differed, but the subjects' answers were similar

on corresponding experiment 1 and experiment 2 pictures, the schema formed in experiments 1 and 2 must be different. There are three places where this difference could occur: in the criteria curves used for feature assessment and scoring, in the relative weighting of the "many" and "multi-axis" features, and in the relative weighting of the ships, subs, and aircraft overall threat features.

If the difference was solely in the feature criteria curves, then the score for the features "many ships", "many submarines", and "many aircraft" assigned to n platforms in the low all-out attack experiments would also be assigned to 1.5 x n platforms in the high all-out attack experiments. Instead, n platforms in the low set got the same score as a x n platforms in high set, with a = 1.22 for ships, 1.28 for air, and 1.34 for submarines. The difference between these numbers may reflect a contribution from the usual notion of "many"; nine ships fits the natural category "many ships" better than does six ships.

These numbers indicate that while the feature criteria curves account for much of the difference, feature combination and feature weighting are also important. In these experiments, the initial feature combination rule, combining "many" with "multi-axis" to yield "overall" for the ships, submarine, and aircraft features made up half of the difference. The rest was made up by subjects' weighting features that received high scores more in the low all-attack cases than in the high all-out attack cases.

The feature criteria curves inferred by the subjects do not replicate the curves in the model used to construct the training and test pictures, particularly in the barrier pictures. Such replication is not expected, of course,

given the amount of information provided in the training about the relative contribution of different factors. In the training the subjects were told only the overall picture quality and the names of the features that are weak or strong. Since they were never given any numerical information relating particular feature characteristics to corresponding feature scores, and since there are many combinations of feature scores and combination rules able to produce each picture value, subjects did not have the information necessary to infer the actual feature scaling curves used to develop the pictures.

Feature weighting. A comparison of the results in the three experiments shows that weighted geometric means of the features predict attack and barrier assessments better than do the unweighted means, that subjects use simple schema-specific rules to attain the weights, and that the feature importance scores reflect the feature's strength (feature assessment score) as well as its weight in the geometric mean.

The rule for attaining weights in the all-out and barrier cases were significantly different. For the all-out attacks, the weights were the feature assessment scores. For the barriers the weights were .75 for the weaker feature and .25 for the stronger one. The existence of simple rules for assigning feature weights avoids a requirement for special criteria curves or complex information processing methods for weight determination. Using such simple rules conserves both memory and information processing resources.

When these experiments were designed it was thought that feature weights would be closely related to subjects' ratings of feature importance. This relationship was observed in the all-out attack; it was not observed in the barrier experiments. In fact, what was observed was that "importance" was a confounding

of two factors: feature assessment score (how characteristic that feature is of a strong attack or barrier) and feature weight. In the all-out attack, these two factors correlated and the weight seemed to be derived from the assessment score. In the barriers, the weighting rule that worked was the one used to develop the training picture. It rated a barrier's strength mostly from its weaker component. For barriers, weighting features by assessment score reduced the correspondence between the weighted geometric mean of features and the assessments of barrier quality.

Feature combination. All three of the experiments tested the extent to which the weighted geometric mean of the feature scores predicted subjects' assessments of attack or barrier quality. In all three cases, the correlation between the attack and barrier quality assessments and weighted feature geometric means, averaged over subjects, exceeded 0.97. In addition, the absolute difference between the weighted means and the quality assessments was very small, averaging about .35 over all experiments.

Conclusions and Further Applications. The proposed schema and information processing model provide an excellent explanation of the subjects' performance in these situation assessment task. The subjects formed the schema from a sequence of examples described in terms of features. Their assessments of the overall situation appeared to be derived from their assessments of the situation features, and these seemed to be based on objective measurable properties of the presented pictures. Further, the schema so formed linked easily with concepts subjects had obtained previous to the training. Subjects used these concepts in their situation evaluations.

While it is not likely that this specific model will be equally useful for understanding every kind of situation assessment task, it is possible that variants will prove useful for a broad range of such tasks. For example, scripts, which have been shown useful for understanding social situations, are a variant of the proposed model. The features of scripts are events and the time relationship among events. Their feature criteria curves address the characteristics of the script events and time arrangements.

It is also possible that the very simple schema presented here will prove to be important building blocks of more elaborate structures requiring a more extensive set of related schema. These structures may have several levels of schema hierarchy, and may include schema composed of more abstract features.

Schema for situation assessment support "intuitive" decision making. This kind of decision making is based on recognizing that an observed situation is similar to other situations in which particular decisions or strategies generally work well. "Intuitive" decision making requires data in memory that supports the necessary similarity assessment. The schema examined in these experiments may play an important role in such assessments.

REFERENCES

Abelson, R. P. (1981). Psychological status of the script concept. American Psychologist, 36, 715-729.

Bower, G. H., Black, J. B. & Turner, T. J. (1979). Scripts in memory for text. Cognitive Psychology, 11, 177-220.

Chase, W. G., & Simon, H. A. (1973). The mind's eye in chess. In W. G. Chase (Ed.), Visual Information Processing. New York: Academic Press, 1973.

diSessa, A. A. (1983). Phenomenology and the evolution of intuition. In D. Gentner & A. Stevens (Eds.), Mental models. Hillsdale, N.J.: Lawrence Erlbaum Assoc.

Elio, R., & Anderson, J. R. (1981). The effects of category generalization and instance similarity on schema abstraction. Journal of Experimental Psychology: Human Learning and Memory, 7, 397-417.

Gati, I., & Tversky, A. (1984). Weighting common and distinctive features in perceptual and conceptual judgements. Cognitive Psychology, 16, 341-370.

Hayes-Roth, B., & Hayes-Roth, F. (1977). Concept learning and the recognition and classification of exemplars. Journal of Verbal Learning and Verbal Behavior, 16, 321-338.

Kahneman, D., & Tversky, A. (1973). On the Psychology of Prediction. Psychological Review, 80, 237-251.

Kintsch, W., & Greeno, J. G. (1985). Understanding and solving word arithmetic problems. Psycological Review, 92 (1), 109-129.

Larkin, J. H. (1983). The role of problem representation in physics. D. Gentner & A. Stevens (Eds.), Mental models. Hillsdale, N.J.: Lawrence Erlbaum Assoc.

Lewis, M. W., & Anderson, J. R. (1985). Discrimination of operator schemata in problem solving: Learning from examples. Cognitive Psychology, 17, 26-65.

Mayer, R. E. (1982). Memory for algebra story problems. Journal of Educational Psychology, 74, 199-216.

Noble, D. & Truelove, J. (1985). Schema-based theory of information presentation for distributed decision making. (Technical Report 4-028-85) Vienna, VA: Engineering Research Associates. (NTIS No. AD A163150).

Pryor, J. B. & Merluzzi, T. V. (1985). The role of expertize in processing social interaction scripts. Journal of Experimental Social Psychology, 21, 362-379.

Raven, J. E. (1958). Standard progressive matrices, sets A, B, C, D, and E. London: H. K. Lewis and Company, Ltd.

Rumelhart, D. E. (1980). Schemata: The building blocks of cognition. In R. Spiro, B. Bruce, & W. Brewer (Eds.), Theoretical issues in reading comprehension. Hillside, NJ: Lawrence Erlbaum Assoc., 33-58.

Rumelhart, D. E. (1981). Understanding understanding. Tech. Rep. CHIP 100, LaJolla, CA: University of CA, San Diego.

Rumelhart, D. E., & Ortony, A. (1977). The representation of knowledge in memory. In R.C. Anderson, R.J. Spiro, & U.E. Montague (Eds.), Schooling and the acquisition of knowledge. Hillsdale, N.J.: Lawrence Erlbaum Assoc.

Smith, E. E., & Medin, D. L. (1981). Categories and concepts. Cambridge, MA: Harvard University Press.

Thorndyke, P. W. (1977). Cognitive structures in comprehension and memory of narrative discourse. Cognitive Psychology, 9, 77-110.

Tversky, A. (1977). Features of similarity. Psychological Review, 84, 327-352.

Tversky, A., & Kahneman, D. (1980). Causal schemas in judgements under uncertainty. In M. Fishbein (Ed.), Progress in social psychology. Hillside, NJ.: Lawrence Erlbaum Assoc.

Tversky, A., & and Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgement. Psychological Review, 90, 293-315.

Tversky, B., & Hemenway, K. (1984). Objects, parts, and categories. Journal of Experimental Psychology, General, 113, 169-193.

Van Dijk, T. A., & Kintsch, W. (1983). Strategies of discourse comprehension. New York: Academic Press.

Zimmermann, H. J., and Zysno, P. (1980). Latent connectives in human decision making. Fuzzy Sets and Systems, 4, 37-51.

APPENDIX A

TABLES FOR DESIGN OF MATERIALS

FEATURE CRITERIA CURVES FOR SHIPS

OBJECTIVE MEASURE Number of Ships	SUBJECTIVE SCORE "Many Ships"
2	2
3	3
4	4 - 5
5 - 6	7

To attain overall strength add 1 or 2 to "many ships" if multi-axis is two.

FEATURE CRITERIA CURVES FOR AIRCRAFT

OBJECTIVE MEASURE Number of Aircraft	SUBJECTIVE SCORE "Many Aircraft"
2	2
3	3
5	5
6 - 9	7
10	8
13 - 15	10

To attain overall strength, subtract 1 or 2 from "many aircraft" if multi-axis is one.

FEATURE CRITERIA CURVES FOR SUBMARINES

NUMBER		SURROUNDEDNESS	
OBJECTIVE MEASURE Number of Submarines	SUBJECTIVE SCORE "Many Submarines"	OBJECTIVE MEASURE Number of Quadrants Covered	SUBJECTIVE SCORE "Multi-axis"
2	2	1	2
3	6	2	6
4	8	3	8
5	9	4	10
6	10		

Overall strength is geometric mean of "many submarines" and "multi-axis."

Table A-1. Construction of test and training pictures for all-out attacks: criteria curve data used for feature scoring

PICTURE NUMBER	FEATURE QUALITY SPREAD	ALL-OUT ATTACK STRENGTH	SHIP STRENGTH OVERALL STRENGTH	SHIP STRENGTH NUMBER OF SHIPS	SHIP STRENGTH NUMBER OF AXIS	AIRCRAFT STRENGTH OVERALL STRENGTH	AIRCRAFT STRENGTH NUMBER OF A/C	AIRCRAFT STRENGTH NUMBER OF AXIS	SUBMARINE STRENGTH OVERALL STRENGTH	SUBMARINE STRENGTH NUMBER OF SUBS	SUBMARINE STRENGTH QUADRANT COVERED
1	L	9	9	7	2	9	13	2	9	6	4
2	ML	8	8	6	1	7	8	1	9	4	3
3	MH	7	7	5	1.5	9	12	1.5	6	3	2
4	HIGH	7	8	6	2	4	4	1	10	6	4
5	HIGH	6	5	4	1	10	15	2	4	2	2
6	ML	5	4	4	1	6	6	1.5	5	4	1
7	MH	4	8	7	2	2	2	1	4	3	1
8	ML	4	3	3	1	3	3	1	7	4	2
9	ML	3	2	2	1	5	5	2	3	2	2
10	L	2	2	2	1	2	2	1	2	3	1

TABLE A-2. Design for Test Pictures for Experiment 1. The design for experiment two is the same, except that each picture has 50% more ships, aircraft and submarines.

FEATURE CRITERIA CURVES FOR BARRIERS

OBJECTIVE LENGTH	SUBJECTIVE f(L)	OBJECTIVE GAP*	SUBJECTIVE g(G)
1"	1	6	1
2	2	5	2
3	3	4	4
4	4	3	5
5	5	2	6
6	7	1	8
7	9	0	10
8	10		

*Add 1.5" to gap to find physical separation between platforms bordering longest internal gap.

SCORE FOR BARRIER EFFECTIVENESS =

$$f(L)^{P} \quad g(G)^{1-P}$$

where $P = .75$ if $f(L) < g(G)$

$P = .25$ otherwise

TABLE A-3. Construction of test and training pictures for barriers; criteria curve data used for feature scoring.

PICTURE	RATING NUMBER ROUND	RATING CALC.	DESCRIPTION NUMBER SHIPS	DESCRIPTION LENGTH	f(L)	GAP SIZE	g(G)	CATEGORY* LENGTH/GAP SIZE	COMMENTS
1	10	10	8	8"	10	-	10	H/H	PROTOTYPE
2	7	6.8	6	5.5"	6	-	10	M/H	
3	2	1.8	2	1.25	1	-	10	L/H	TRAINING AND
4	8	7.6	6	8"	10	1.5"	7	H/M	TEST
5	2	1.8	2	8"	10	6.5"	1	H/L	
6	6	5.6	4	5"	5	1"	8	M/M	
7	4	3.7	2	6"	7	4.5"	3	M/L	
8	3	3.3	2	2.5"	2.5	1"	8	L/M	TRAINING
9	5	4.9	3	7"	9	4"	4	H/L	ONLY
10	7	6.5	5	6.5"	8	2"	6	H/M	
11	3	3	3	2"	2	-	10	L/H	
12	5	5.21	3	4.75"	5	2"	6	M/M	TEST
13	6	5.6	4	6.5"	8	3.5"	5	H/M	ONLY
14	9	8.4	7	6.5"	8	-	10	H/H	
15	4	3.566	3	3"	3	1"	6	L/M	
16	2	(same as #5)						H/L	
17	8	(same as #4)						H/M	
18	10	(same as #1)						H/H	
19	2	(same as #3)						L/H	
20	7	(same as #2)						M/H	

*FOR LENGTH		FOR GAP SIZE	
H	6-8"	H	None
M	4-6"	M	1-3.5"
L	2-4"	L	4-6.5"

TABLE A-4a. Design for test and training pictures in experiment 3.

ISLANDS AND OFF-CENTERS

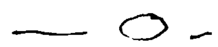
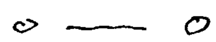
PICTURE	FUNCTIONS AS	LOOKS LIKE	DESCRIPTION	COMPARED IN PICTURES FUNCTIONS AS	COMPARED IN PICTURES LOOKS LIKE	
21	#2	#12		35	29	
22	#2	#11		37	34	ISLANDS
23	#4	# 5		28	31	
24	#14	#13		33	30	
25	#1	# 3		32	36	
26	#11	# 2		39	38	
27	#3	#12		41	40	OFFSETS

TABLE A-4b. Design of test and training pictures for experiment 3. Islands, Barriers, and Off-Centers.

OFFICE OF NAVAL RESEARCH

Engineering Psychology Program

TECHNICAL REPORTS DISTRIBUTION LIST

Dr. Carl Alluire
Office of the Deputy Under
Secretary of Defense
OUSDRE (E&LS)
Pentagon, Room 3D129
Washington, DC 20301

Engineering Psychology Program
Office of Naval Research
Code 1142EP
800 North Quincy Street
Arlington, VA 22217-5000 (3 copies)

Dr. Charles Holland
Office of Naval Research
Code 1133
800 North Quincy Street
Arlington, VA 22217-5000

J. Randy Simpson
Statistics Program Code 1111SP
Office of Naval Research
800 North Quincy Street
Arlington, VA 22217-5000

Information Sciences Division
Code 1113
Office of Naval Research
800 North Quincy Street
Arlington, VA 22217-5000

James G. Smith
Code 1211
Office of Naval Research
800 North Quincy Street
Arlington, VA 22217-5000

Dr. Randall P. Schumaker
NRL A. I. Center
Code 7510
Naval Research Laboratory
Washington, DC 20375-5000

Special Assistant for Marine Corps
Matters
Code 00MC
Office of Naval Research
800 North Quincy Street
Arlington, VA 22217-5000

CDR James Offutt
Office of the Secretary of Defense
Strategic Defense Initiative Organization
Washington, DC 20301-7100

Director
Technical Information Division
Code 2627
Naval Research Laboratory
Washington, DC 20375-5000

Dr. Michael Melich
Communications Sciences Division
Code 7500
Naval Research Laboratory
Washington, DC 20375-5000

Mr. J. S. Lawson, Jr.
4773-C Kahala Avenue
Honolulu, HI 96816

Naval Training Equipment Center
ATTN: Technical Library
Orlando, FL 32813

Human Factors Department
Code N-71
Naval Training Equipment Center
Orlando, FL 32813

Mr. Norm Beck
Combat Control System Department
Code 35
Naval Underwater Systems Center
Newport, RI 02840

Human Factors Engineering
Code 441
Naval Oceans Systems Center
San Diego, CA 92152

Dr. Gary Poock
Operations Research Department
Naval Postgraduate School
Monterey, CA 93940

Mr. H. Talkington
Engineering & Computer Science
Code 09
Naval Ocean Systems Center
San Diego, CA 92152

Dr. A. L. Slafkosky
Scientific Advisor
Commandant of the Marine Corps
Code RD-1
Washington, DC 20380

Dr. L. Chmura
Computer Sciences & Systems
Code 7592
Naval Research Laboratory
Washington, DC 20375-5000

Dr. Michael Letsky
Office of the Chief of Naval
Operations (OP-01B7)
Washington, DC 20350

Human Factors Engineering
Office of Naval Technology
Code 0722
800 North Quincy Street
Arlington, VA 22217-5000

Mr. Philip Andrews
Naval Sea Systems Command
NAVSEA 61R
Washington, DC 20362

Aircrew Systems Branch
System Engineering Test Directorate
U.S. Naval Test Center
Patuxent River, MD 20670

Dr. George Moeller
Human Factors Engineering Branch
Naval Submarine Base
Submarine Medical Research Lab.
Groton, CT 06340

Dr. Robert Blanchard
Code 17
Naval Personnel Research and
Development Center
San Diego, CA 92152-6800

LCDR T. Singer
Human Factors Engineering Division
Naval Air Development Center
Warminster, PA 18974

LT Dennis McBride
Human Factors Branch
Pacific Missile Test Center
Point Mugu, CA 93042

Dr. Kenneth L. Davis
Code 1114
Office of Naval Research
800 North Quincy Street
Arlington, VA 22217-5000

LCDR R. Carter
Office of Chief of Naval Operations
(OP-933D3)
Washington, DC 20350-2000

CAPT M. Moroney
Naval Air Development Center
Code 602
Warminster, PA 18974

Dr. Harry Crisp
Code N51
Combat Systems Department
Naval Surface Weapons Center
Dahlgren, VA 22448

Human Factors Branch
Code 3152
Naval Weapons Center
China Lake, CA 93555

Dr. Rabinder N. Madan
Code 1114SE
Office of Naval Research
800 North Quincy Street
Arlington, VA 22217-5000

Dr. Eugene E. Gloye
ONR Detachment
1030 East Green Street
Pasadena, CA 91106-2485

Dr. Glen Allgaier
Artificial Intelligence Branch
Code 444
Naval Electronics Ocean System Center
San Diego, CA 92152

Dr. Sherman Gee
Command and Control Technology,
(MAT 0721)
Office of Naval Technology
800 North Quincy Street
Arlington, VA 22217-5000

Dr. Robert A. Fleming
Human Factors Support Group
Naval Personnel Research &
Development Center
1411 South Fern Street
Arlington, VA 22202

Dr. Edgar M. Johnson
Technical Director
U.S. Army Research Institute
Alexandria, VA 22333-5600

Technical Director
U.S. Army Human Engineering Lab.
Aberdeen Proving Ground, MD 21005

Director, Organizations and Systems
Research Laboratory
U.S. Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333-5600

Dr. M. Katz
Director, Basic Research
Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333-5600

Dr. A. Fregly
U.S. Air Force Office of Scientific
Research
Life Science Directorate, NL
Bolling Air Force Base
Washington, DC 20332-6448

Dr. R. K. Dismukes
Associate Director for Life Sciences
AFSOR
Bolling AFB
Washington, DC 20032-6448

Dr. M. C. Montemerlo
Information Sciences & Human
Factors Code RC
NASA HQS
Washington, DC 20546

Dr. Alan Leshner
Deputy Division Director
Division of Behavioral and Neural
Sciences
National Science Foundation
1800 G Street, N.W.
Washington, DC 20550

Dr. Clinton Kelly
Defense Advanced Research Projects Agency
1400 Wilson Boulevard
Arlington, VA 22209

Dr. Jesse Orlansky
Institute for Defense Analyses
1801 N. Beauregard Street
Alexandria, VA 22311

Dr. Stanley Deutsch
NAS-National Research Council (COHF)
2101 Constitution Avenue, N.W.
Washington, DC 20418

Dr. Deborah Boehm-Davis
Department of Psychology
George Mason University
4400 University Drive
Fairfax, VA 22030

Dr. Paul E. Lehner
PAR Technology Corp.
7926 Jones Branch Drive
Suite 170
McLean, VA 22102

Dr. James H. Howard, Jr.
Department of Psychology
Catholic University
Washington, DC 20064

Dr. Robert Wherry
Analytics, Inc.
2500 Maryland Road
Willow Grove, PA 19090

Dr. Edward R. Jones
Chief, Human Factors Engineering
McDonnell-Douglas Astronautics Co.
St. Louis Division
Box 516
St. Louis, MO 63166

Dr. Lola L. Lopes
Department of Psychology
University of Wisconsin
Madison, WI 53706

Dr. Joseph G. Wohl
Alphatech, Inc.
3 New England Executive Park
Burlington, MA 10803

Dr. Marvin Cohen
Decision Science Consortium, Inc.
Suite 721
7700 Leesburg Pike
Falls Church, VA 22043

Dr. Andrew P. Sage
Assoc. V.P. for Academic Affairs
George Mason University
4400 University Drive
Fairfax, VA 22030

Dr. Richard Pew
Bolt Beranek & Newman, Inc.
50 Moulton Street
Cambridge, MA 02238

Dr. Hillel Einhorn
Graduate School of Business
University of Chicago
1101 E. 58th Street
Chicago, ILL 60637

Dr. John Payne
Graduate School of Business
Administration
Duke University
Durham, NC 27706

Dr. Azad Madni
Perceptronics, Inc.
6271 Variel Avenue
Woodland Hills, CA 91364

Dr. David Castanon
ALPHATECH, Inc.
111 Middlesex Turnpike
Burlington, MA 01803

Dr. E. Douglas Jensen
Carnegie-Mellon University
Computer Science Department
Pittsburgh, PA 15213

Dr. David Noble
Engineering Research Associates
8618 Westwood Center Drive
Vienna, VA 22180

Dr. Bruce Hamill
The Johns Hopkins University
Applied Physics Lab
Laurel, MD 20707

Dr. A. Ephremides
University of Maryland
Electrical Engineering Department
College Park, MD 20742

Professor Michael Sovereign
Joint Command, Control &
Communications Curriculum
Code 74
Naval Postgraduate School
Monterey, CA 93943

Dr. David L. Kleinman
Electrical Engineering & Computer
Science Department
University of Connecticut
Storrs, CT 06268

Mr. Robert L. Stewart
The Johns Hopkins University
Applied Physics Laboratory
Laurel, MD 20707

Dr. Alexander Levis
Massachusetts Institute of Technology
Lab Information & Decision Systems
Cambridge, MA 02139

Dr. Michael Athans
Massachusetts Inst. of Technology
Lab Information & Decision Systems
Cambridge, MA 02139

Dr. Leonard Adelman
PAR Technology Corp.
7962 Jones Branch Drive
Suite 170
McLean, VA 22102

Dr. P. Papantoni-Kazakos
University of Connecticut
Department of Electrical Engineering
& Computer Science (U-157)
Storrs, CT 06268

Professor Wayne F. Stark
University of Michigan
Department of Electrical Engineering
& Computer Science
Ann Arbor, MI 48109

Dr. Kepi Wu
Space and Naval Warfare Systems
Code 611
Washington, DC 20363-5100

Defense Technical Information Center
Cameron Station, Building 5
Alexandria, VA 22314 (2 copies)